

3. FORMES LINEAIRES VERSUS FORMES NON LINEAIRES : UNE APPLICATION AU CHOIX MODAL TRAIN/AVION

Jusqu'à récemment, la plupart des études empiriques utilisait des données en préférences révélées (ou "revealed preference data" ou RP). Ces données sont issues de l'observation de situations réelles. Dans un contexte de choix modal, elles concernent les niveaux actuels des attributs de chaque mode, les caractéristiques socio-économiques et le choix modal des usagers. Depuis, sous l'impulsion d'un courant d'origine anglo-saxonne, le type de données utilisées s'est élargi aux données en préférences déclarées (ou "stated preference data" ou SP). Relevant de la méthodologie de l'évaluation contingente, ces dernières, fréquemment employées dans le domaine du marketing, décrivent le comportement des individus confrontés à des situations hypothétiques.

La combinaison des avancées théoriques proposées dans la deuxième section et des différentes méthodes de recueil de données structure nos investigations empiriques. Ces dernières s'appuient sur un échantillon de 500 usagers du train et de l'avion, pour lesquels nous disposons après enquête en site propre⁹, des caractéristiques de réseau du mode choisi et du mode alternatif (prix, temps d'attente, de rabattement,...) et de certaines caractéristiques socio-économiques (PCS, revenu,...). Notre travail procède selon la séquence suivante :

- une première étape consiste en l'analyse des choix révélés des usagers. Différentes formes fonctionnelles pour les fonctions d'utilité indirecte, allant de la forme linéaire standard jusqu'aux formes non linéaires plus élaborées que nous avons proposées en amont, sont estimées. Ces dernières, outre une meilleure qualité statistique, débouche sur l'identification d'une fonction de valeurs du temps révélées, alternative intéressante à l'unicité des valeurs fournies par les modèles linéaires standards ;

- dans une deuxième étape, les techniques des préférences déclarées sont appliquées à la partie hypothétique du questionnaire. Il s'agit ici, de déterminer une fonction de valeurs du temps déclarées applicables à des zones de validité étendues par rapport à la situation révélée.

⁹ Cette enquête a été réalisée par nos soins grâce à l'appui de la Direction Régionale Languedoc-Roussillon de la SNCF et de la CCI de Montpellier, responsable de l'aéroport Montpellier-Méditerranée. Elle porte sur une période allant du 6 janvier au 4 février 1997. Une première présélection a limité la population échantillonnée aux usagers concernés par un trajet pour lequel il existe un équivalent dans le mode alternatif. Ainsi, les lignes aériennes considérées englobent toutes les lignes intérieures au départ de Montpellier et les zones ferroviaires couvertes sont les zones SNCF 7 et 9.

31. Approche révélée

311. Formes linéaires standards

Les modèles estimés sont les suivants :

Modèle 1 : $V_i = a - \lambda pt_i - k tt_i$, $i = 1(\text{train}), 2(\text{avion})$ ¹⁰

avec - $tt_i = tach_i + t_i$ ¹¹ où $tach_i$ et t_i correspondent respectivement aux temps d'acheminement (temps d'accès et de rabattement) et en véhicule associés au mode i ;

- pt_i désigne le prix "porte à porte" du mode i .

Modèle 2 : $V_i = a - \lambda pt_i - k_i tt_i$, $i = 1,2$.

Le modèle 2 permet de réintroduire la spécificité des coefficients temporels, grâce à l'utilisation de coefficients k_i différenciés selon le mode.

Les résultats obtenus sont les suivants :

Modèle 1¹² :

Discrete Choice Model

Maximum Likelihood Estimates

Log-Likelihood..... -303.8865
Restricted (Slopes=0) Log-L. -316.0751
Chi-Squared (3)..... 24.37723
Significance Level..... 0.2083529E-04

N[0,1] used for significance levels.

Variable	Coefficient	Std. Error	t-ratio	Prob t
CSTE	-0.52794	0.2214	-2.384	0.01711
PT	-0.94101E-03	0.2601E-03	-3.618	0.00030
TT	-0.94253E-03	0.7265E-03	-1.297	0.19451

¹⁰ Par la suite, l'indice 1 désignera systématiquement le train et l'indice 2 l'avion.

¹¹ Dès que l'on passe à un niveau appliqué, il convient de considérer le problème de la désagrégation du temps de trajet. En effet, l'intuition suggère une valorisation différenciée selon que le temps correspond à un temps d'acheminement, d'attente ou en véhicule. Sur notre échantillon, seuls les temps en véhicule présentent des ordres de grandeur pour lesquels l'approximation linéaire nous a semblé discutable. Aussi, nous avons opté dans notre analyse pour une agrégation des temps d'acheminement et du temps en véhicule pour les termes d'ordre 1. En ce qui concerne les termes d'ordre supérieur, seuls les temps en véhicule sont conservés. De plus, le temps d'attente a été écarté en raison de sa non-significativité systématique quels que soient les modèles estimés.

¹² L'estimation est réalisée grâce au logiciel LIMDEP version 6.0 et 7.0 d'Econometric Software.

Modèle 2 :

Discrete Choice Model

Maximum Likelihood Estimates

Log-Likelihood..... -300.4910

Restricted (Slopes=0) Log-L. -316.0751

Chi-Squared (4)..... 31.16826

Significance Level..... 0.2828093E-05

N[0,1] used for significance levels.

Variable	Coefficient	Std. Error	t-ratio	Prob t
CSTE	-1.2328	0.3545	-3.478	0.00051
PT	-0.77511E-03	0.2702E-03	-2.868	0.00413
TT1	-0.90329E-03	0.7341E-03	-1.230	0.21854
TT2	-0.56173E-02	0.1972E-02	-2.849	0.00438

Ils permettent d'élaborer les tableaux synoptiques suivants :

TABLEAU 1 : ANALYSE STATISTIQUE COMPARATIVE DES MODELES 1 ET 2

	Significativité des coefficients	Test du Khi2	¹³ Pseudo R^2 et ¹⁴ Pseudo \bar{R}^2
Modèle 1	Bonne	$H_0 : \lambda = k = 0$ rejetée	$R^2 = 0.039$ $\bar{R}^2 = 0.032$
Modèle 2	Bonne	$H_0 : \lambda = k_1 = k_2 = 0$ rejetée	$R^2 = 0.049$ $\bar{R}^2 = 0.04$

¹³ La mesure utilisée est celle proposée par McFadden [15] : $\rho^2 = 1 - \frac{\ell(\beta)}{\ell(0)}$ où $\ell(0)$ représente la vraisemblance du modèle contraint, dans lequel tous les coefficients hormis la constante sont assujettis à la nullité, et où $\ell(\beta)$ désigne la vraisemblance du modèle non contraint.

¹⁴ Le pseudo R^2 ajusté fait référence à la statistique d'Akaike : $\bar{\rho}^2 = 1 - \frac{\ell(\beta) - k}{\ell(0)}$ où k représente le nombre de degrés de liberté, c'est-à-dire le nombre de paramètres hors constante estimés.

TABEAU 2 : ANALYSE ECONOMIQUE COMPARATIVE DES MODELES 1 ET 2

			Modèle 1	Modèle 2
Elasticités directes *	- Coût du trajet	Train	-0.231	-0.190
		Avion	-0.363	-0.301
	- Temps en véhicule	Train	-0.187	-0.178
		Avion	-0.034	-0.203
Elasticités croisées *	- Coût du trajet	Train	0.488	0.400
		Avion	0.172	0.143
	- Temps en véhicule	Train	0.045	0.269
		Avion	0.139	0.134
VTTS (frs/h)	Train	60	70	
	Avion	60	435	

* Les élasticités sont calculées aux points moyens de l'échantillon révélé.

Malgré les bons résultats des tests de significativité des coefficients, des tests du Khi 2 et l'obtention d'ordre de grandeurs réalistes pour les valeurs du temps et les élasticités, ces deux modèles pâtissent de pseudo R² faibles. Ces statistiques, appréciant la part de la variance expliquée par la régression, il semblerait donc qu'une partie importante du comportement de choix modal ne soit pas pris en compte par une modélisation linéaire. Ce résultat, plutôt négatif, renvoie à l'estimation de formes fonctionnelles plus complexes.

322. Formes non linéaires

L'adaptation à notre étude de la forme théorique suggérée à la section 2 (relation 6) débouche sur une utilité représentative du type :

$$V_i = cste - \tilde{\alpha} pt_i + \frac{\beta}{2}(pt_i^2 - 2R pt_i) + \gamma(Rt_i - pt_i t_i) + \tilde{\delta}t_i - \frac{\xi}{2}t_i^2$$

Pour des raisons techniques, nous posons dans l'estimation :

$$x = pt_i^2 - 2R pt_i ; \quad rmp_t = Rt_i - pt_i t_i \quad \text{et} \quad st_i = t_i^2$$

L'estimation du modèle 3 donne :

Modèle 3 :

Discrete Choice Model

Maximum Likelihood Estimates

Log-Likelihood..... -241.2041

Restricted (Slopes=0) Log-L. -316.0751

Chi-Squared (8)..... 149.7421

Significance Level..... 0.0000000

N[0,1] used for significance levels.

Variable	Coefficient	Std. Error	t-ratio	Prob t
CSTE	-0.15533	0.6123	-0.254	0.79974
PT	-0.36042E-02	0.5767E-03	-6.249	0.00000
X	-0.83076E-07	0.1906E-07	-4.359	0.00001
RMPT	-0.23982E-06	0.8152E-07	-2.942	0.00326
TT1	-0.61993E-02	0.2489E-02	-2.491	0.01273
ST1	0.95625E-05	0.3091E-05	3.094	0.00198
TT2	-0.79258E-02	0.2765E-02	-2.867	0.00414
ST2	0.33723E-04	0.1823E-04	1.850	0.06431

Les tests et résultats associés à ce modèle sont décrits dans le tableau ci-dessous :

TABLEAU 3 : ANALYSE DU MODELE 3

Tests de signification des coefficients		H_0 : "nullité du coefficient étudié" rejetée	
Test du Khi2		H_0 : "nullité simultanée de tous les coefficients sauf constante" rejetée	
Pseudo R^2 et Pseudo \bar{R}^2		$R^2 = 0.237$ et $\bar{R}^2 = 0.214$	
Fonctions de VTTS		$VTTS_i = \frac{\partial V_i / \partial t_i}{\partial V_i / \partial pt_i}$	
Elasticités directes*	- Coût du trajet	Train	-0.362
		Avion	-0.511
	- Temps en véhicule	Train	-0.583
		Avion	-0.167
	- Revenu	Train	-1.190
		Avion	0.722
Elasticités croisées*	- Coût du trajet	Train	0.843
		Avion	0.219
	- Temps en véhicule	Train	0.276
		Avion	0.354

* Les élasticités sont calculées aux points moyens de l'échantillon révélé.

L'utilisation d'une forme non linéaire accroît le pouvoir explicatif du modèle (passage du pseudo R^2 de 0.049 à 0.237. et du pseudo \bar{R}^2 de 0.04 à 0.214). De plus, on remarque que les élasticités calculées pour des valeurs moyennes identiques sont largement supérieures en valeur absolue à celles du cas linéaire. En ce point moyen, le modèle linéaire sous estime donc les réactions des usagers. D'une manière générale si l'on fait varier continûment chacune des variables on constate que le modèle linéaire conduit à une sur ou sous estimation systématique des effets marginaux.

Les fonctions de VTTS fournissent maintenant des valeurs du temps différenciées selon le mode, le revenu, les temps de trajet en véhicule et les prix pratiqués. A titre d'exemple, nous donnons pour chacun des modes, les tableaux suivants relatifs au revenu moyen et à différentes combinaisons prix/temps¹⁵.

TABLEAU 4 : VALEURS DU TEMPS EN TRAIN POUR UN REVENU DE 13470 FRS (FRS/H)

t1\pt1	180	210	240	270	300	330	360	390	420	450	480	510
3h30	239	238	237	236	235	234	232	231	230	229	228	227
3h50	223	222	221	220	219	218	216	215	214	213	212	211
4h10	207	206	205	204	202	201	200	199	198	197	196	195
4h30	190	189	188	187	186	185	184	183	182	181	180	179
4h50	174	173	172	171	170	169	168	167	166	165	164	163
5h10	157	156	155	154	153	152	152	151	150	149	148	147
5h30	140	139	138	138	137	136	135	134	133	133	132	131
5h50	123	122	122	121	120	119	118	118	117	116	115	115
6h10	106	105	105	104	103	102	102	101	100	100	99	98
6h30	89	88	87	87	86	86	85	84	84	83	82	82
6h50	71	71	70	70	69	68	68	67	67	66	66	65

TABLEAU 5 : VALEURS DU TEMPS EN AVION POUR UN REVENU DE 13470 FRS (FRS/H)

t2\pt2	470	550	630	710	790	870	950	1030	1110	1190	1270	1350
0h50	321	318	314	310	307	303	300	296	293	290	287	284
1h00	294	290	287	283	280	277	274	270	267	264	261	258
1h10	266	262	259	256	253	250	247	244	241	238	236	233
1h20	238	235	232	229	226	223	221	218	215	213	210	208
1h30	210	207	204	202	199	197	194	192	189	187	184	182
1h40	182	179	177	174	172	170	167	165	163	161	159	157
1h50	153	151	149	147	145	143	141	139	137	135	133	131
2h00	125	123	121	119	117	116	114	112	110	109	107	105

¹⁵ Les revenus moyens et les combinaisons prix/temps retenus renvoient aux valeurs observées sur notre échantillon.

Il faut être circonspect quant à l'interprétation des valeurs figurant dans ces tableaux. Bien évidemment, les valeurs associées à des combinaisons prix/temps particulières "ne doivent pas être prises au pied de la lettre.". Il convient plutôt de retenir des valeurs plus parlantes du type VTTS aux valeurs moyennes et/ou médianes. Notons par ailleurs que les VTTS aux points moyens, ombrées dans les tableaux, sont différentes des valeurs du modèle linéaire.

De manière générale, quel que soit le mode, on observe à tarif fixé (respectivement à temps fixé), une VTTS décroissante par rapport au temps (respectivement par rapport au prix). Cependant, ces résultats qui présentent la VTTS comme inversement reliée au coût et à la longueur du trajet sont peut-être imputables aux caractéristiques de notre échantillon. D'un point de vue théorique, le sens de cette relation est donné par le signe des dérivées partielles de la VTTS par rapport au prix et au temps. Après examen de celles-ci, il nous paraît impossible de conclure de façon catégorique à l'émergence d'un signe particulier¹⁶.

Rappelons de plus que les résultats bruts obtenus ci-dessus seraient bien évidemment améliorés par une meilleure prise en compte de l'hétérogénéité de l'échantillon, qui passe par la recherche de critères de segmentation pertinents. Enfin, pour en finir avec l'approche révélée, il convient de mentionner un problème statistique passé sous silence jusqu'ici : la présence de colinéarité entre les variables, leurs carrés et leurs interactions. Sans être capables de "chiffrer" les répercussions de cette anomalie il est néanmoins évident qu'elle n'est pas sans incidence sur les estimations obtenues.

32. Approche déclarée

321. Présentation liminaire

Jusqu'au milieu des années quatre vingt, l'intérêt des théoriciens, en économie des transports, s'est essentiellement porté sur le développement de modèles de comportements révélés de choix discret. Depuis, les techniques des préférences déclarées, jusqu'alors cantonnées à d'autres domaines¹⁷, ont connu un vif essor. Les principales raisons de ce succès proviennent de la capacité de ces méthodes :

- à s'appliquer à des situations de choix non observables sur le marché (alternatives et/ou attributs virtuels) ;

¹⁶ Soulignons que la relation décroissante entre le temps et la VTTS observée sur notre échantillon, émerge de façon récurrente dans la littérature (Cf. : Hensher[11]). Il n'en va pas de même en ce qui concerne la relation VTTS/prix.

- à fixer le niveau des attributs de façon à améliorer la qualité statistique des estimateurs par rapport aux situations révélées.

L'expérience hypothétique peut revêtir les deux formes suivantes :

- l'individu interrogé classe l'ensemble des alternatives qui lui sont proposées. Ce classement peut être ordinal ou cardinal¹⁸ ;

- l'individu interrogé choisit une seule alternative dans l'ensemble de choix qui lui est proposé.

Ce deuxième cas de figure s'impose comme le pendant évident du cadre d'analyse révélé. Il débouche naturellement sur une formalisation en terme de choix discret déclaré estimable à l'aide de modèles logit. C'est celui que nous retenons par la suite.

La mise en place de l'expérience passe par la définition des attributs, de leurs niveaux et des ensembles de choix soumis aux personnes interrogées. Chacune de ces étapes est sous le contrôle exclusif de l'expérimentateur qui est amené à opérer un arbitrage entre la qualité statistique et la simplicité du modèle. Concrètement, après avoir sélectionné les attributs pertinents pour l'expérience, le modélisateur choisit le niveau des attributs adapté à sa problématique. Chaque combinaison de niveaux d'attributs correspond alors à une alternative possible. Malheureusement, l'ensemble de ces alternatives, que nous appellerons par la suite "Full Factorial Design", devient très vite ingérable (Exemple : une expérience avec seulement 4 attributs à 3 niveaux chacun génère $3^4=81$ combinaisons). Néanmoins, c'est la seule structure qui garantisse l'indépendance statistique de l'estimation des effets de chaque attribut sur les réponses. L'analyste est donc amené à construire un ensemble de choix réduit, ou "Fractional Factorial Design" au prix cependant d'une perte d'efficacité statistique. Afin de limiter cette perte d'efficacité, il convient de construire le "Fractional Factorial Design" selon une méthodologie particulière¹⁹.

322. *Expérience*

Initialement, l'objectif de cette expérience était double. Il s'agissait dans un premier temps, d'estimer une fonction de valeurs du temps déclarées pour des zones de validité élargies par rapport aux situations révélées. Dans un second temps, l'introduction dans l'expérience d'un prix environnemental, attribut virtuel géré par nos soins, devait permettre de

¹⁷ Marketing, et psychométrie.

¹⁸ Dans la littérature anglo-saxonne, ces métriques sont respectivement désignées par les expressions "ranked data" et "rated data".

¹⁹ Cf. : Louviere [13] et Hahn et Shapiro [9].

déterminer les arbitrages opérés par les usagers dans un contexte hypothétique d'intériorisation de la pollution. Il n'y a pas lieu, dans cet article, de s'étendre sur ce dernier point, aussi nous ne commentons pas les résultats relatifs aux prix environnementaux.

Formellement, un individu interrogé est amené à choisir entre un train et un avion. Chaque mode proposé est alors décrit par l'ensemble des caractéristiques suivant :

- un prix "économique" qui correspond au prix d'un billet SNCF ou d'une compagnie aérienne tel qu'il pourrait lui être actuellement proposé ;
- un prix "environnemental" correspondant à un montant supplémentaire qu'il devrait acquitter pour couvrir le coût de la pollution qu'il génère en voyageant ;
- une durée de trajet comprenant les temps en véhicule et de correspondance.

Pour chacun des attributs, trois niveaux ont été retenus. Ils correspondent, pour le temps et le prix "économique", à des niveaux moyens²⁰ réels, effectivement observés. Pour le prix "environnemental", les valeurs finalement retenues sont celles prônées par le rapport Boiteux²¹ à un facteur multiplicatif près²². Le tableau suivant présente ces valeurs :

TABLEAU 6 : NIVEAUX DES ATTRIBUTS

	TRAIN			AVION		
Prix "éco"	285	415	545	700	1000	1300
Prix "envir"	2	4	6	70	120	170
Temps	300	495	815	50	100	200

Le nombre de combinaisons possibles des différents niveaux d'attributs détermine la structure de l'expérience. Avec trois attributs à trois niveaux pour chacun des deux modes (soit six caractéristiques), le nombre maximum de combinaisons est de $3^6=729$. On réduit alors ce "Full Factorial Design" à 27 ensembles de choix. Le "Fractional Factorial Design" ainsi obtenu représente le nombre minimum d'ensembles de choix avec des propriétés statistiques acceptables. Il est présenté dans le tableau ci-dessous :

²⁰ Ils sont obtenus après comptabilisation des différentes O-D nationales à 400, 700 et 1000 km.

²¹ Cf. : Boiteux [4]

²² Les valeurs proposées dans le rapport sont de 0.12 et 3.6 cts/km/voyageur. Les niveaux absolus pour les distances retenues sont alors de 0.48, 0.84, et 1.2 frs/voyageur pour le train et 14.4, 25.2 et 36 frs/voyageur pour l'avion. Nous avons jugé ces valeurs trop faibles pour un contexte de choix hypothétique. Aussi, avons nous finalement conservé les ordres de grandeur relatifs, tout en multipliant globalement les niveaux absolus par cinq.

TABLEAU 7 : PRESENTATION DU "FRACTIONAL FACTORIAL DESIGN" DE L'EXPERIENCE

	TRAIN			AVION		
	Prix "éco"	Prix "envir"	Temps	Prix "éco"	Prix "envir"	Temps
A(1)	285	2	5h	700	70	0h50
B(2)	285	4	5h	1000	170	1h40
C(3)	285	6	5h	1300	120	3h20
A(4)	285	4	8h15	700	120	3h20
B(5)	285	6	8h15	1000	70	0h50
C(6)	285	2	8h15	1300	170	1h40
A(7)	285	6	13h35	700	170	1h40
B(8)	285	2	13h35	1000	120	3h20
C(9)	285	4	13h35	1300	70	0h50
A(10)	415	4	5h	700	120	1h40
B(11)	415	6	5h	1000	70	3h20
C(12)	415	2	5h	1300	170	0h50
A(13)	415	6	8h15	700	170	0h50
B(14)	415	2	8h15	1000	120	1h40
C(15)	415	4	8h15	1300	70	3h20
A(16)	415	2	13h35	700	70	3h20
B(17)	415	4	13h35	1000	170	0h50
C(18)	415	6	13h35	1300	120	1h40
A(19)	545	6	5h	700	170	3h20
B(20)	545	2	5h	1000	120	0h50
C(21)	545	4	5h	1300	70	1h40
A(22)	545	2	8h15	700	70	1h40
B(23)	545	4	8h15	1000	170	3h20
C(24)	545	6	8h15	1300	120	0h50
A(25)	545	4	13h35	700	120	0h50
B(26)	545	6	13h35	1000	70	1h40
C(27)	545	2	13h35	1300	170	3h20

Pour simplifier la tâche des enquêteurs, seuls trois ensembles de choix sont proposés à chacun des individus interrogés. L'utilisateur doit alors, dans les trois situations proposées, choisir le mode qu'il préfère.

Du point de vue statistique, pour comprendre comment la forme analytique de la fonction d'utilité indirecte déclarée est dérivée de l'expérience, il faut rappeler que l'analyse conjointe est fondée sur le principe selon lequel on peut décomposer l'effet des attributs sur le choix :

- en effet direct, on parle alors d'effet principal (ou "main effect") et on traduit cette influence par une relation linéaire ;

- et/ou en effet indirect, et s'opérer via des interactions du deuxième ordre ou d'ordre supérieur.

Des attributs à trois niveaux ne contiennent de l'information que jusqu'aux termes quadratiques et débouchent sur une forme analytique similaire à celle que nous proposons dans la section 2. Tout naturellement, nous avons donc pensé que l'approche déclarée pouvait constituer un nouveau domaine de validation de notre forme analytique. Elle nous permet également d'étendre les zones de validité par rapport à l'enquête révélée.

323. Résultats

Nous estimons sur données déclarées une forme similaire à celle du modèle 3²³. Dans le modèle 4, seuls les paramètres significatifs sont conservés :

Modèle 4 :

```
Discrete Choice Model
Maximum Likelihood Estimates
Log-Likelihood..... -399.9478
Restricted (Slopes=0) Log-L. -505.3043
Chi-Squared ( 7)..... 210.7130
Significance Level..... 0.0000000
N[0,1] used for significance levels.
```

Variable	Coefficient	Std. Error	t-ratio	Prob t
CSTE	2.6388	0.9252	2.852	0.00434
PT	-0.34839E-02	0.4231E-03	-8.234	0.00000
X	-0.70952E-07	0.1324E-07	-5.358	0.00000
TT1	-0.13726E-01	0.3496E-02	-3.926	0.00009
ST1	0.89811E-05	0.2997E-05	2.997	0.00273
PE1	0.88297E-01	0.5322E-01	1.659	0.09711
TT2	-0.54147E-02	0.1438E-02	-3.765	0.00017

Les tests et résultats associés à ce modèle sont décrits dans le tableau suivant :

²³ L'estimation est effectuée sur données réelles. La présence résiduelle de corrélation entre le niveau des attributs réels, imputable au passage du "Full Factorial Design" au "Fractional Factorial Design", pourrait être neutralisée par l'utilisation de codes orthogonaux polytomiques (Cf. : Louvière [13]). Cependant, la nécessité de conserver le revenu comme variable explicative nous a interdit ce recours.

TABLEAU 8 : ANALYSE DU MODELE 4

Tests de signification des coefficients	<i>H₀ : "nullité du coefficient étudié" rejetée</i>		
Test du Khi2	<i>H₀ : "nullité simultanée de tous les coefficients sauf constante" rejetée</i>		
Pseudo R² et Pseudo \bar{R}^2	R ² =0.209 et \bar{R}^2 =0.197		
Fonctions de VTTS	$VTTS_i = \frac{\partial V_i / \partial t_i}{\partial V_i / \partial pt_i}$		
Elasticités directes*	- Coût du trajet	Train	-0.428
		Avion	-1.361
	- Temps en véhicule	Train	-0.964
		Avion	-0.355
	- Revenu	Train	-0.308
		Avion	0.392
Elasticités croisées*	- Coût du trajet	Train	1.069
		Avion	0.546
	- Temps en véhicule	Train	0.279
		Avion	1.227
* Les élasticités sont calculées aux points moyens de l'échantillon déclaré.			

Les fonctions de valeurs du temps issues du modèle 4 sont présentées dans les tableaux ci-dessous :

TABLEAU 9 : VALEURS DU TEMPS EN TRAIN POUR UN REVENU DE 8400FRS (FRS/H)

t1/pt1	285	315	345	375	405	435	465	495	525	555
5h00	214	214	214	213	213	213	212	212	211	211
6h00	187	186	186	186	185	185	185	184	184	184
7h00	159	159	158	158	158	158	157	157	157	156
8h00	131	131	131	131	130	130	130	130	129	129
9h00	104	103	103	103	103	103	102	102	102	102
10h00	76	76	76	75	75	75	75	75	75	75
11h00	48	48	48	48	48	48	48	48	47	47
12h00	20	20	20	20	20	20	20	20	20	20

TABLEAU 10 : VALEURS DU TEMPS EN AVION POUR UN REVENU DE 8400 FRS (FRS/H)

t2\pt2	700	770	840	910	980	1050	1120	1190	1260	1330
0h50	136	135	135	134	134	133	133	132	131	131
1h10	136	135	135	134	134	133	133	132	131	131
1h30	136	135	135	134	134	133	133	132	131	131
1h50	136	135	135	134	134	133	133	132	131	131
2h10	136	135	135	134	134	133	133	132	131	131
2h30	136	135	135	134	134	133	133	132	131	131
2h50	136	135	135	134	134	133	133	132	131	131
3h10	136	135	135	134	134	133	133	132	131	131
3h30	136	135	135	134	134	133	133	132	131	131

La prudence dans l'interprétation prônée pour l'approche révélée reste évidemment de mise. Pour permettre encore une fois une comparaison directe avec les formes linéaires, les valeurs du temps associées aux niveaux moyens des attributs sur notre échantillon sont ombrées dans les tableaux. L'expérience déclarée conduit à des valeurs du temps en train systématiquement supérieures aux valeurs obtenues via l'enquête révélée. Le phénomène inverse est observé pour l'avion. De même on constate une surévaluation des élasticités temps du train par rapport au cas révélé. L'asymétrie Train/Avion provient d'un défaut de construction de l'échantillon déclaré restreint aux seuls usagers effectifs du train. La non concordance des valeurs révélées et déclarées s'explique en partie par la présence de biais inhérents à toute expérience hypothétique (justification du choix réel à posteriori, biais stratégique,...). Une méthode d'estimation combinée RP/SP atténuerait sans aucun doute l'effet de ces biais.

En ce qui concerne le sens des relations entre les VTTS et les variables prix, temps et revenu, on observe les phénomènes suivants :

- VTTS croissantes par rapport au revenu ;
- VTTS décroissantes par rapport au temps et au prix.

Ces relations sont semblables à celles du cas révélé. Cependant, lorsque nous réalisons l'estimation sur données codées²⁴, le sens de la liaison entre prix et VTTS s'inverse. Rappelons toutefois que pour donner plus de poids à ce résultat, il faudrait mener une expérience avec endogénéisation du revenu.

D'une manière générale, les résultats obtenus sur données déclarées semble être un élément supplémentaire en faveur de formes polynomiales plus complexes que celle en terme de coût généralisé.

4. CONCLUSION

Plus que dans l'obtention de valeurs numériques particulières, l'intérêt de ce travail réside davantage dans l'exposé de la méthodologie à suivre pour dériver des fonctions de valeurs du temps véritablement comportementales, c'est-à-dire, issues d'une approche intégrée programme allocatif/choix discret. Au niveau pratique, cette étude a le mérite d'éclairer les distorsions dans l'affectation des ressources susceptibles d'être générées par une utilisation abusive des formes linéaires, et ouvre la voie à des politiques de transport plus adaptées. A un niveau heuristique, la méthodologie mise en évidence peut, de plus, être transposée à la valorisation de tout autre attribut.

Parallèlement, cette recherche nous a permis de dépasser le cadre de l'analyse révélée en nous donnant l'occasion de nous essayer aux nouvelles techniques de validation empirique offertes par l'approche déclarée. La complémentarité évidente de ces deux types d'approches a conduit à l'élaboration d'une méthode d'estimation combinée RP/SP (cf. : Morikawa [17]) permettant de bénéficier de leurs avantages respectifs. Cette phase ultime sera l'objet de nos travaux à venir.

²⁴ On rappelle ici que seul l'usage de codes polytomiques permet de circonscrire totalement la présence résiduelle de colinéarité et que la structure de notre expérience nous autorise ce codage uniquement sur un modèle sans revenu.

REFERENCES BIBLIOGRAPHIQUES

- [1] Bates J.J. (1987), « Measuring Travel Time Values with a Discrete Choice model : A Note », *Economic Journal*, volume n°97, pages 493-498.
- [2] Ben-Akiva M. et Morikawa T. (1990), « Estimation of Switching Models from Revealed Preferences and Stated Intention », *Transportation Research*, volume n°24, pages 485-495.
- [3] Ben-Akiva M., Morikawa T., Shiroishi F. (1991), « Analysis of the Reliability of Preference Ranking Data », *Journal of Business Research*, volume n°23, pages 253-268.
- [4] Boiteux M. (sous la direction) (1994), *Transports : Pour un meilleur choix des investissements*, La documentation française.
- [5] Bouyaux P. (1988), « Modélisation désagrégée de la demande de transport urbain : une application à la ville de Rennes », *Revue d'Economie Régionale et Urbaine*, volume n°5, pages 783-810.
- [6] Bruzelius N. (1979), *The Value of Travel Time*, London, Croom Helm.
- [7] De Serpa A.J. (1971), « A Theory of the Economics of Time », *Economic Journal*, volume n°81, pages 828-845.
- [8] Domenich T. A., et McFadden D. (1975), *Urban Travel Demand : A behavioural Analysis*, North-Holland, Amsterdam.
- [9] Hahn G.J., et Shapiro S.S. (1966), « A Catalog and Computer Program for the Design and Analysis of Orthogonal Symmetric and Asymmetric Fractional Factorial Experiments », Technical Report 66-C165, Schenectady, NY : *General Electric Research and Development Center* .
- [10] Hensher D. (1994), « Stated Preference Analysis of Travel Choices : The State of Practice », *Transportation*, volume n°21, pages 107-135.
- [11] Hensher D. (1995), « Value of travel time savings in personal and commercial automobile travel », *Institute of Transport Studies*, ITS-WP-95-4.
- [12] Hivert L., Orfeuil J-P, et Troulay P. (1988), « Modèles désagrégés de choix modal : réflexions méthodologiques autour d'une prévision de trafic », Rapport *INRETS* n°67.

- [13] **Louviere J.J.** (1991). *Analyzing Decision Making*, Sage Publications, n°67.
- [14] **Louviere J.J. et Hensher D.** (1996), « Stated Preference and Choice Methods for Land Use and Transportation Planning Applications », Short Course, June 24-27, *Center for Urban Studies School of Urban and Public Affairs*, Portland State University.
- [15] **McFadden D.** (1973), « Conditionnal Logit Analysis of Qualitative Choice Behaviour », dans *Frontiers in Econometrics*, Ed. Zarembka, Academic Press, pages 105-142.
- [16] **McFadden D.** (1981), « Econometric Models of Probabilistic Choice », dans Manski C.F., and McFadden D. (eds.), *Structural Analysis of Discrete Data*, MIT Press. Cambridge, pages 198-272.
- [17] **Morikawa T.** (1994), « Correcting State Dependence and Serial Correlation in the RP/SP Combined Estimation Method », *Transportation*, volume n°21, pages 153-165.
- [18] **MVA Consultancy, Institute of Transport Studies University of Leeds and Transport Studies Unit University of Oxford** (1987), *The Value of Travel Time Savings*, Policy Journals , London, England.
- [19] **Truong T.P. et Hensher D.** (1985), « Measurement of Travel Times Values and Opportunity Cost from a Discrete-Choice Model », *Economic Journal*, Volume n°95, pages 438-451.

Questions relatives aux exposés de J. Polack (CTS-Londres), D. Lemaire (EURO-TUNNEL) d'une part, T. Blayac et A. Causse (LAMETA) :

J. Bravais

D.Lemaire ne nous a pas dit de quelle façon il a traité la question du prix dans son modèle de préférence déclarée. Je crois que c'était sous forme de pourcentage.

A. Causse

Nous avons joué sur trois attributs d'après la construction des designs de l'analyse déclarée, nous avons choisi 3 niveaux pour chacun de ces attributs qui correspondent aux niveaux réels des lignes à 300 km, 700 km, et 1000 km. Nous avons ensuite construit ce design selon les techniques statistiques propres à l'analyse déclarée. Nous avons donc considéré les tarifs économiques réels affectés en moyenne aux parcours de 300 km.

J. Bravais

Vous avez fait une moyenne sur l'ensemble des lignes TGV ! Pourquoi n'avez-vous pas plutôt pris des tarifs existants sur les lignes en question ?

A. Causse

Parce que nous les avions en révélé.

J. Bravais

Le but est tout de même de poser des questions aux gens qui portent sur un référentiel qui est le leur pour un voyage donné. Sinon, la réponse que peuvent fournir les gens n'a pas de valeur.

A. Causse

C'est justement le principe de la préférence déclarée que de pousser les gens à répondre à des situations qu'ils ne vivent pas forcément. Le but de l'enquête étant de déterminer leur opinion en matière de coût de mesures liées à l'environnement, nous n'avions pas besoin de connaître leur situation réelle, nous l'avions en révélé. Les questions posées concernent les points suivants: le trajet qu'ils viennent de faire, le temps mis pour se rendre à la gare, le temps du trajet en train, le temps mis pour se rendre ensuite à leur destination finale, le coût au niveau du temps d'acheminement et du temps de parcours, le coût comparatif du même trajet en avion (s'il est connu) et la durée de ce trajet. Nous sommes ainsi parvenus à cibler toutes les situations effectives. A partir de ce modèle révélé, nous avons mené une modélisation logique. Ce qui nous intéressait ensuite était de placer la personne dans la situation hypothétique de tel trajet à effectuer dans tel type de train ou d'avion, et de savoir quel moyen de transport elle choisissait alors.

J. Bravais

Les réponses à des situations hypothétiques n'en restent pas moins elles-mêmes hypothétiques!

A. Causse

Bien sûr, c'est pourquoi nous essayons à présent de combiner la source révélée et la source déclarée à partir de méthodes d'estimation dont J.Polack nous a parlé.

J. Bravais

Je pense qu'il faut être très vigilant quant aux questions que l'on pose aux gens. Il ne faut pas poser des questions auxquelles nous aimerions avoir une réponse, mais des questions auxquelles ils sont susceptibles de répondre. Il vaut peut-être alors mieux se baser sur une méthode totalement statistique qui ne se fonde pas sur du recueil d'informations auprès des gens. Je pense que la probabilité sur de telles questions que les personnes interrogées agissent comme elles prétendraient le faire dans vos questionnaires est extrêmement faible.

D. Lemaire

Dans notre cas, c'est à dire celui de la préférence déclarée, je pense qu'il est essentiel de commencer par administrer aux enquêtés des questions leur permettant de mieux se repositionner dans la mémoire du vécu du voyage qu'ils ont fait cet été, par exemple, avant de pratiquer des combinaisons de modes d'enquêtes. Cela les resitue dans un contexte de voyages, et c'est à ce moment là qu'il faut les faire basculer dans un jeu. Il faut toujours travailler par rapport à un référentiel. Au moment de l'enquête, le SHUTTLE ou l'EUROSTAR ne faisait pas forcément encore partie de l'environnement culturel du transport. Il y a effectivement un biais dans ce cas.

A. Causse

Nous avons également des référentiels précis, mais étant donnée la construction du design, vous ne pouvez pas toujours développer les attributs qui correspondent au dernier voyage de la personne.

M. Papillon

Une question pour A.Causse : dans la première technique, vous faites référence au test du Chi 2 ? satisfaisant ?

D. Lemaire

Il s'agit d'un aspect non linéaire, je pense qu'il faut donc considérer la question statistique avec prudence.

A. Causse

D'autant plus que nous avons travaillé sur l'échantillon brut avec les seules variables qui étaient justifiées au niveau économique. Nous n'avons pas du tout joué sur la segmentation de la population, et il est probable que si nous introduisions des données telles que l'âge, le sexe, la fréquence, nous améliorerions considérablement ce modèle. Le niveau obtenu est à mon avis raisonnable pour l'utilisation de seulement trois caractéristiques.

M. Papillon

M. Lemaire, pouvez-vous nous donner un aperçu des résultats issus de votre enquête de type préférence déclarée ?

D. Lemaire

La question qui se pose est de savoir si la préférence déclarée mène à une situation de transition entre une situation révélée et une situation à long terme et quel sera le temps nécessaire pour que ces résultats soient valides. Nos analyses ont malheureusement été perturbées par l'incendie qui s'est déclaré dans le tunnel et qui a bouleversé toutes les prévisions de montée en charge.

M. Houée

M. Polack, vous n'avez pas abordé une des difficultés auxquelles Christian Calzada et moi-même avons été confrontés au cours de l'étude multimodale de la longue distance (l'aérien et le routier) qui est la définition des modalités binaires de chaque attribut de telle façon que l'un des attributs ne capte pas la quasi-totalité de l'explication ? Bien que nous ayons essayé d'obtenir une amplitude suffisante pour les modalités des autres attributs, l'attribut « type d'offre multimodale » s'avère de loin le plus discriminant. Que recommandez-vous dans ce cas, étant donné que la taille des échantillons n'est pas adaptée au type d'analyse interactive qu'ont développée récemment Martin Lee-Gosselin, et Bruno Faivre d'Arcier en France ? Pensez-vous qu'il soit nécessaire de tester le questionnaire lui-même afin de s'assurer de la meilleure façon de déterminer l'importance relative de chaque attribut en les confrontant les uns aux autres, ou de procéder à d'autres types d'amélioration ?

J. Polack

Le problème que vous rencontrez là est très courant dans ce type d'enquête. Les spécifications modales paraissent incomplètes; tout ce qui pourrait aider à développer d'autres facteurs permettant d'améliorer le pouvoir explicatif du modèle serait utile. Il faudrait pour ce faire développer des travaux préliminaires qualitatifs, basés sur de petits échantillons dans le but de déterminer les attributs réels.

Le problème réside bien sûr dans l'application du modèle; il y a une distinction entre le modèle d'estimation et le modèle appliqué à des prévisions. Par exemple les résultats de ces travaux qualitatifs sont que si certains de ces attributs comme le niveau de congestion paraissent importants dans l'explication des choix des gens, on peut les inclure dans le questionnaire SP. Mais ceci n'est utile dans l'optique prévisionnelle que si l'on est capable de prévoir l'évolution de ce niveau de congestion, ce qui est plutôt délicat.

Un autre aspect auquel, si je le comprends bien, votre problème est relié, tient à ce que lorsque l'on pourvoit le répondant en information dans le cadre de l'exercice SP, cette information est supposée parfaite. Or les choix que les gens font dans la vie courante sont souvent imparfaits. Quand on utilise un modèle SP dans le but de faire des prévisions sur le réel, il faut prendre en compte ces imperfections. Par exemple, si il y a des imperfections systématiques dans la perception des différents attributs, comme le temps ou le prix, ils ne seront pas pris en compte dans le questionnaire SP, car l'information fournie au gens est parfaite. Si on applique le modèle en utilisant les valeurs objectives de ces quantités pour établir des prévisions, ces prévisions ne seront plus fonction de ces imperfections de perception. Il faut donc modéliser ces erreurs de perception comme faisant partie de la technique de conception du questionnaire; c'est pourquoi l'élaboration d'un questionnaire SP peut se révéler très complexe, contrairement à celle d'un questionnaire RP. La méthode SP nécessite donc de réintroduire systématiquement les données réelles dans le processus d'enquête à un moment donné.

Session 6 :

**A SOURCE D'ENQUETE UNIQUE, AXES D'ANALYSE MULTIPLES :
L'EXEMPLE DE L'ETC**

** Analyse neuronale des voyages à longue distance*
C. CALZADA (SES), S. BAYOMOG (*Université Paris I*)

** Segmentation des déplacements et modélisation
du partage modal à longue distance*
G. DUMARTIN, E. MAURICE (DR)

** Modèles d'interactions spatiales et recherche de partitionnement*
C. CALZADA (SES), C. GRASLAND (*Equipe PARIS, Université Paris I*)

***Les voyages à longue distance dans l'ETC 93-94:
une approche neuronale***

S. BAYOMOG (SAMOS Université Paris I)
C. CALZADA (SES/DEE)

Le point fort des enquêtes ménages INSEE dites structurelles réalisées en 1966-67, 1973-74, 1981-82, et 1993-94, cette dernière faisant l'objet de cette étude (enquête transports et communications (ETC)), c'est leur caractère multimodal et de fournir une vision d'ensemble cohérente des habitudes et pratiques de voyages des ménages, en particulier à longue distance.

Le dépouillement et l'exploitation des informations recueillies se heurte aux problèmes classiques de données incomplètes (non-réponses partielles et totales) et de synthèse intelligente du matériau collecté. Ceci à des fins de politiques de transports et de prévisions et dans un souci de rationalisation des dispositifs de systèmes d'informations, qui demain se devront d'être nécessairement plus légers et plus efficaces, dans le but de centrer au mieux le recueil d'information sur des cibles de populations a priori pertinentes.

Les méthodes classiques d'analyses des données telles que les analyses canoniques et factorielles ont prouvé depuis longtemps leur efficacité à traiter des problèmes de corrélations entre variables au sein d'une population. Malgré tout celles-ci présentent quelques inconvénients: elles sont intrinsèquement linéaires et ne permettent pas de représenter graphiquement de manière satisfaisante les corrélations entres plus de deux variables croisées.

C'est pourquoi des méthodes dites neuronales ont été développées, elles présentent plusieurs avantages:

- elles constituent tout d'abord un outil d'analyse exploratoire des données (EDA), complémentaire à d'autres outils, ne nécessitant pas la construction d'hypothèses a priori, hypothèses de fait contraignantes sur la structure du modèle à tester,
- un autre avantage est de permettre de 'révéler' des relations comportementales dites du second ordre. La nature exploratoire de ces techniques permet en cela d'identifier des variables que nous qualifierons de complexes (styles et cycles de vie), dont on sait qu'elles peuvent conditionner des comportements spécifiques de transport et sur lesquelles il s'avère impossible de questionner directement mais dont on peut appréhender les contours (proxy) à l'aide des variables socio-démographiques usuelles,
- elles permettent la prise en compte de phénomènes de nature intrinsèquement non linéaires, plus proches des notions de cycles de vie, de chaînes d'activité, etc.
- enfin contrairement aux techniques classiques, les algorithmes utilisés (Kohonen en particulier) ne nécessitent pas de calculs matriciels importants.

Parmi les algorithmes neuronaux, celui d'auto-organisation dit de Kohonen est particulièrement célèbre, il permet en particulier d'obtenir une carte plane de données multidimensionnelles nombreuses. A cette fin et afin de construire une 'typologie intelligente' croisant caractéristiques socio-démographiques des individus se déplaçant et voyages effectués à plus de 100 km*, une plate-forme informatique a été construite pour la mise en oeuvre de l'algorithme KACM (M. Cottrell, S. Ibbou, 1995) sous le logiciel SAS®.

* voyage: ensemble de parcours effectués par une personne (dite B) à plus de 100 km du domicile. Un voyage est constitué au minimum d'un aller-retour c'est à dire de 2 déplacements.

Typologies des voyages à longue distance

Christian CALZADA, Samuel BAYOMOG (SAMOS Université Paris I)

Le point fort des enquêtes structurelles ménages transports Insee, réalisées en 1966-67, 1973-74, 1981-82, et 1993-94 (enquête transports et communications (ETC)), c'est leur caractère multimodal et leur capacité à fournir une vision d'ensemble cohérente des habitudes et pratiques de voyages des ménages, en particulier à longue distance.

Le dépouillement et l'exploitation des informations recueillies se heurtent aux problèmes classiques de données incomplètes (non-réponses partielles ou totales) et de synthèse intelligente du matériau collecté (construction de typologies). Les méthodes classiques d'analyses multidimensionnelles des données telles que les analyses canoniques et factorielles ont prouvé depuis longtemps leur efficacité à traiter des problèmes de corrélations entre variables au sein d'une population. Cependant ces approches sont intrinsèquement linéaires et ne permettent pas de représenter graphiquement de manière satisfaisante les corrélations entre plus de deux variables croisées. C'est pourquoi d'autres méthodes dites neuronales ont été développées. La présente note analyse les principes de ces méthodes et leur apport dans l'exploitation de l'enquête ménages transports.

Description des données utilisées

Sont analysés ci-après les voyages au sens de l'ETC 1993-94, autrement dit l'ensemble des parcours effectués par une personne (dite B) à plus de 100 kilomètres du domicile. Un voyage est constitué au minimum d'un aller-retour, c'est à dire de deux déplacements.

Un voyage est décrit par les attributs suivants : le motif principal associé, le mode d'hébergement principal, sa durée (en nuitées), le mode de transport principal emprunté, la distance parcourue pendant le voyage, l'origine-destination. En outre pour chaque individu voyageant on dispose des indicateurs socio-démographiques suivants : sexe, âge, condition familiale..., niveau d'études, PCS, strate de résidence, ... etc...; ainsi que de variables dites *d'opportunité de voyage* comme l'existence de contacts privés ou professionnels, la prise en charge ou non des frais de déplacements, parc automobile du ménage, etc....

L'analyse porte sur un « échantillon utile » de 18 499 voyages, soit 80% de l'échantillon exhaustif des individus B qui décrivent leurs opportunités de contacts, sur la période allant de mai 1993 à avril 1994, dans un questionnaire rétrospectif de leurs déplacements à longue distance (à plus de 80 km à vol d'oiseau du lieu de domicile) et qui remplissent un questionnaire sur leurs voyages pendant les trois mois qui précèdent la visite de l'enquêteur.

	<i>Nombre de voyages</i>	<i>Nombre d'individus</i>
0 voyage	0	6 145
entre 1 et 6 voyages	14 563	7 203
+ de 6 voyages	8 499	720
<i>total</i>	23 062	14 068
<i>échantillon étudié</i>	18 499	-

Les méthodes classiques d'analyses des données

Plusieurs analyses multidimensionnelles ont été menées de front afin de dégager différentes typologies permettant de prendre en compte l'ensemble des types de variables descriptives des voyages : attributs du voyage, intensité de mobilité, caractéristiques socio-démographiques des individus mobiles, variables dites d'opportunité de voyage.

Typologie globale des individus effectuant des voyages à longue distance

Cette première approche vise à l'analyse des liaisons entre variables décrivant les voyages (variables dites 'actives'), données caractéristiques des individus mobiles ainsi que variables dites 'sociales' (variables dites 'supplémentaires'); elle est menée au moyen d'une Analyse des Correspondances Multiples (ACM) classique* et appelle les commentaires généraux suivants :

- le nombre de voitures particulières (parc) varie dans le même sens que le nombre de voyages,
- deux types de pratique des voyages s'opposent : les personnes qui déclarent pratiquer régulièrement la conduite d'une voiture et ceux à pratique occasionnelle ou nulle,
- le nombre de voyages varie en sens inverse de la distance parcourue pour un voyage, autrement dit plus on voyage, moins on va loin,
- une opposition entre deux grandes zones d'origine des voyages : l'Ile de France et le reste de la France.

Malgré tout si l'on souhaite disposer d'une vue synthétique de l'ensemble des informations disponibles, la projection des 94 modalités des variables étudiées, comme c'est le cas de l'analyse développée ci-haut [*], rend les graphiques vite illisibles.

Apport des méthodes dites neuronales

Les méthodes neuronales présentent quelques avantages :

- contrairement aux techniques classiques, les algorithmes utilisés (Kohonen en particulier) ne nécessitent pas de calculs matriciels importants. Parmi les algorithmes neuronaux, celui d'auto-organisation dit de Kohonen est particulièrement célèbre [voir description en annexe], il permet en particulier d'obtenir une carte plane lisible de données multidimensionnelles nombreuses,
- elles constituent tout d'abord un outil d'analyse exploratoire des données (EDA), complémentaire à d'autres outils, ne nécessitant pas la construction d'hypothèses contraignantes a priori sur la structure du modèle à tester,
- un autre avantage est de permettre de 'révéler' des relations comportementales dites du second ordre. La nature exploratoire de ces techniques permet en cela d'identifier des variables que nous qualifierons de complexes (styles et cycles de vie), dont on sait qu'elles peuvent conditionner des comportements spécifiques de transport et sur lesquelles il s'avère impossible de questionner directement mais dont on peut appréhender les contours à l'aide des variables socio-démographiques usuelles,
- enfin elles permettent la prise en compte de phénomènes de nature intrinsèquement non linéaires, plus proches des notions de cycles de vie, de chaînes d'activité, etc....

Typologie neuronale des individus effectuant des voyages à longue distance

Dans cette étape on a procédé à l'analyse simultanée des correspondances entre toutes les variables entrant dans l'étude. Vue l'hétérogénéité des variables analysées cette démarche ne semble pas opératoire mais ce travail a pour but de rassembler dans une même étape la description des types de voyages et la typologie des individus qui les effectuent, elle permet de dégager des indices de résumés des voyages et de la constitution des individus qui les réalisent. On étudie les relations entre toutes les variables au moyen de l'algorithme de Kohonen (KACM), analyse suivie d'un regroupement par classification hiérarchique.

Sur la carte de Kohonen [carte1], cinq classes de voyages apparaissent, il s'agit :

- des *voyages d'une journée*, qui sont surtout des voyages domicile travail et sont entrepris par des personnes travaillant en Ile de France, en provenance du bassin parisien. Les personnes qui effectuent ce type de voyage sont des ouvriers, de niveau d'étude CAP ou BEP, elles vivent en couple avec enfant et réalisent au moins 6 voyages dans l'année.
- des *voyages à moyenne distance*, effectués par des individus appartenant à des ménages possédant au moins une voiture, faisant au moins deux voyages en voiture et résidant soit dans l'Ouest de la France dans des communes de plus de 100 000 habitants soit dans la région parisienne. Dans ce groupe on trouve également des couples non mariés, âgés de 26 à 40 ans qui conduisent régulièrement et sont de profession intermédiaire, des hommes ou des membres de ménages avec au moins trois enfants, ainsi que des femmes de niveau d'étude baccalauréat et des employés. Les voyages s'effectuent principalement dans la région du sud-ouest ou à destination du bassin parisien.
- des *voyages à une distance comprise entre 500 et 1000 kilomètres* qui se passent généralement chez des amis et durent de une à trois nuitées. Ils sont le fait d'individus disposant d'une résidence secondaire ; de classe d'âge 51-60 ans, de couples mariés ou de ménages sans enfant.
- des *voyages de plus de 1000 kilomètres* qui sont en général effectués par des retraités en avion et durent au moins quatre nuitées, à l'étranger en location. Ces voyages ont en majorité pour origine l'Ile de France (unité urbaine de Paris). Le motif de ce type de voyages est d'ordre privé, ce sont des personnes déclarant avoir plusieurs contacts privés.
- des *voyages pour motif professionnel* qui correspondent à des voyages de personnes qui bénéficient d'une prise en charge (complète ou partielle) et qui déclarent avoir des contacts privés et professionnels, ils sont dans la classe d'âge de 40 à 50 ans, artisans, commerçants, chefs d'entreprise ou cadres supérieurs.

Identification des types de voyages

Chaque voyage est décrit par un certain nombre de variables dites 'principales' : motif, mode d'hébergement, durée, moyen de transport, distance parcourue. On réalise ci-après une classification des modalités de ces variables à l'aide de l'algorithme KACM. La carte n°2 permet d'identifier trois grands groupes de voyages :

- les *voyages à très longue distance* à plus de 1000 kilomètres du domicile, en mode d'hébergement locatif, par avion et de durée supérieure ou égale à quatre nuitées. Ces voyages sont essentiellement d'ordre privé.
- les *voyages à moyenne distance* sur une distance comprise entre 300 et 1000 kilomètres de la zone de résidence chez des parents ou amis, en général en voiture et pour une durée d'une à trois nuitées. Ils sont pour une grande partie d'ordre privé mais peuvent avoir un motif professionnel.
- les *voyages à faible distance* d'une durée qui n'excède pas la journée, généralement en train. Le motif 'principal' de ces voyages est professionnel et souvent de type domicile-travail.

La prise en compte de l'intensité de mobilité

Dans cette partie on incorpore aux variables décrivant chaque voyage, une variable résumant l'intensité de la mobilité (nombre de voyages annuels effectués par l'individu concerné). A cet effet chaque voyage entrant dans l'analyse a été associé à un numéro d'ordre qui indique que l'individu ayant décrit ce voyage a effectué un nombre de voyages au moins égal à ce numéro. On applique ensuite l'algorithme KACM sur l'ensemble des variables décrivant un voyage augmenté de la variable numéro d'ordre du voyage et on obtient la carte n°3.

On retrouve les trois grands groupes de voyages identifiés dans la partie précédente, la plus grande intensité de mobilité (5^{ème} voyage) étant associée aux distances les moins longues (< 300 km).

Eléments de méthodologie de l'algorithme de KOHONEN

L'algorithme d'auto-organisation de Kohonen [T. Kohonen, 1982, 1989, 1995] s'inscrit dans le cadre des méthodes neuronales. De nombreux algorithmes, dérivés de l'algorithme initial, ont été appliqués à des problèmes d'analyse multidimensionnelle des données tels que la classification et l'analyse factorielle simple. L'algorithme KACM employé ci-dessus a été proposé par ** et vise à l'Analyse des Correspondances Multiples via l'algorithme de Kohonen..

Description sommaire

On travaille une matrice 'K' de même dimension que le tableau dit de Burt 'B' mais dont chaque élément est égal à l'élément correspondant de 'B' normalisé par la marge de sa ligne.

Ainsi dans l'algorithme :

- les inputs sont les lignes de la matrice K,
- la distance utilisée est celle du Chi-deux,
- la loi de tirage de l'input μ est donnée par le profil marginal lignes de la matrice K,
- à la dernière itération, chaque input est associé à une unité du réseau, ce qui donne un 1^{er} classement des entrées. Chaque classe correspond à une unité, cette dernière est représentée par le vecteur code qui lui est associé.

Lorsque les poids se stabilisent (dernière itération) chaque unité est représentée par un vecteur code ; on considère ainsi le tableau de données dont les unités constituent les individus décrits par les vecteurs codes. De cette façon le tableau en sortie se prête à tout type d'analyse statistique classique, notamment on peut entreprendre un travail de classification des unités du réseau en groupes homogènes par un indice de similarité de vecteurs codes. Ici le regroupement en 'superclasses' est obtenu par une classification ascendante hiérarchique des unités gagnantes du réseau avec pour indice de similarité celui d'augmentation du critère d'inertie de Ward.

Références Bibliographiques

- ** *M. Cottrell, P. Letrémy : Classification et analyse des correspondances au moyen de l'algorithme de Kohonen: application à l'étude de données socio-économiques. Prépublication du SAMOS n°42 (1995), Université Paris I.*
- *S. Ibou, M. Cottrell : Multiple Correspondence analysis of crosstabulations matrix using the Kohonen algorithm. Prépublication du SAMOS n°49 (1995), Université Paris I.*
- *F. Blayo, M. Verleysen : Les réseaux de neurones artificiels, Que sais-je ?, n° 3042, PUF.*
- *D. Shmuelli, I. Salomon, D. Sheffer : Neural network analysis of travel behaviour: evaluating tools for prediction, Transportation Research - C, Vol. 4. No. 3, pp. 151-166, 1996.*
- *C. Calzada, S. Midenet : Des neurones dans les réseaux, Notes de Synthèse de l'OEST, N°97, décembre 1995.*
- *N. Gouider : Les français voyagent de plus en plus, Notes de Synthèse du SES, N°113, septembre-octobre 1997.*
- *N. Gouider, A. Kerihuel : Les voyages touristiques de courte durée en 1994, Notes de Synthèse du SES, N°114, novembre-décembre 1997.*

Toutes les analyses ont été menées sous le logiciel SAS® version 6.11 (Orlando).

Carte n° 1 : Représentation des modalités sur une carte de Kohonen
 Algorithme KMCA sur 22 variables (400 itérations)

S.O. vie en couple n.d. PCS	Contacts nuls	Couple non marié Age 26-40 ans Conduit régulièr. 2 voitures Prof. intermédiaires Homme Comm. - 20 000 h. Couple avec 2 enf.	Dest. Bassin Par.	0 déplacement Employé		Pers. Seule Orig. Ile de France U.U. de Paris Pers. Seule (ménage)	0 voiture
Age < 25 ans S.O. prat. conduite	BEPC Couple >= 3 enfants	300 < ... < 500 km Dest. Ouest Pas abonnement 1 voiture >= 3 voitures pas de résidence sec. 2ème voyage 3ème voyage voiture Origine Ouest Comm. rurales (RP 90) Comm. + 100 000 h.	Dest. Sud-Ouest Orig. Sud-Ouest	Bao Femme	> 1000 km Dest. Etranger Location . Avion 4 nuits et plus	Abonnement	Conduite occasionnelle Jamais conduite
Origine NPdC		Couple marié Dest. Centre Orig. Centre	Age 51-60 ans Couple sans enfant	500 <...< 1000 km Chez parents/amis Résidence second. 1 à 3 nuits	1er voyage Motif Privé Contacts privés	S.O. p.e.c. frais tr.	Retraité (Prof. Int. Sup.)
Dest. NPdC					Age > 60 ans CEP, DFEO, ss dip. Retraité (Agric., Art.)		Autres modes
		Agric. Exploitants	Dest. Méditerranée Orig. Méditerranée	Comm. 20 000 <...< 100 000 h.			
entièrement p.e.c. frais Motif Professionnel Contacts privés & prof.				Dest. Ile de France		4 ^{ème} et 5 ^{ème} voy.	Etudes sup. au Bac
prof. intellect. sup		Pas de p.e.c. frais		Domicile-Travail Mode Train	< 300 km S.O. hébergement 0 nuits Orig. Bassin Parisien		
Age 41-50 ans partiellement p.e.c. frais	Art., Commerçants, C.E.			5ème voyage et +	CAP-BEP Ouvrier	Couple avec 1 enfant	Dest. Est Orig. Est

294

**Carte n° 2 : Représentation des modalités par l'algorithme KMCA
dans l'analyse des correspondances des voyages**

(120 itérations)

Motif Professionnel	Mode 'Autres'	Train	< 300 km Domicile Travail 0 nuitées
Location Avion		300 < ... < 500 km	
> 1000 km			500 < ... < 1000 km
4 nuitées ou +	Chez parents/amis	Motif Privé Voiture	1 à 3 nuitées

Carte n° 3 : Représentation des variables de voyages et du nombre de voyages

5 ^{ème} voyage et + Domicile-travail	< 300 km 0 nuitées		3 ^{ème} voyage
Train		Motif Professionnel	
	300 < ... < 500 km 2 ^{ème} voyage		500 < ... < 1000 km Chez parents/amis 1 ^{er} voyage Motif Privé Voiture 1 à 3 nuitées
		Autres modes	
4 ^{ème} et 5 ^{ème} voyage			Avion > 1000 km Location 4 nuits ou +

Questions relatives à l'exposé de C. Calzada (SES) et S. Bayomog (Université Paris 1) :

J. Bravais

Allez-vous publier les résultats de ces travaux, notamment la segmentation dans sa totalité ?

C. Calzada

Il y a eu une prépublication au SAMOS, même si un rapport provisoire a été fait au SES depuis plus d'un an.

J. Bravais

Donnez-vous des poids aux différents segments que vous identifiez ?

C. Calzada

Il y a des tableaux de fréquences liés à chacun des segments.

S. Bayomog

A partir de cette méthodologie, on peut arriver à identifier les individus. Ici, nous avons fait une analyse sur le tableau de Burt où la loi de tirage des individus est la marge du tableau. Cela se fait couramment lorsqu'on traite des données avec des variables continues.

C. Calzada

Une précision : pour faire ces exercices croisant variables socio-démographiques tirées de la personne B de l'enquête transport et questionnaire de déplacements de la personne B, sur 4 mois de travail, il y a 3 mois d'apurement. Il y a énormément de données manquantes, et S. Bayomog a fait là un travail considérable de liaison entre ces deux bases d'échantillonnage avant de se lancer dans l'aspect purement statistique.

P. Barette

Vous dites que l'analyse neuronale se rapproche de l'analyse à correspondances multiples; qu'apporte-t-elle en plus ?

S. Bayomog

Je dirais plutôt que ces deux méthodologies sont complémentaires; un petit avantage que nous avons ici est peut-être la représentation de 94 modalités; essayez de les représenter sur l'axe 1-2 ... ce n'est pas facile !

C. Calzada

En effet, l'aspect de visualisation est ici immédiat; on obtient tout de suite une carte de données multidimensionnelles plane, alors que sur une ACM, il y a le choix des axes à additionner, etc ...

S. Bayomog

Et cela ne coûte pas cher ! Du point de vue du temps de traitement, par exemple pour l'obtention de ce tableau 8x8, cela fait 400 intégrations. Pour une ACM, lorsqu'il y a 20.000 individus, l'ordinateur doit être vraiment puissant.

Segmentation des déplacements et modélisation du partage modal à longue distance

G. DUMARTIN (DR/Sous-Direction de la Planification et du Budget)

E. MAURICE (Université Paris X - Nanterre)

Résumé : Des facteurs de natures très diverses concourent au choix d'un mode de transport pour effectuer un déplacement, notamment de longue distance.

L'enquête « transports » (volet « longue distance », déplacements de 100 km et plus) permet d'appréhender ceux qui sont liés :

- * à la nature des déplacements (ex : longueur, motif, période) ;
- * aux caractéristiques des voyageurs (catégorie socioprofessionnelle, personne motorisée ou pas, type de lieu de résidence).

L'application des techniques de l'analyse de données au fichier descriptif des déplacements (issu de l'enquête) a fait apparaître, parmi ces nombreux facteurs, les plus déterminants en terme de choix du mode de transport utilisé.

En particulier, la technique de segmentation de déterminer des catégories de déplacements, les plus homogènes possibles de ce point de vue.

Mais pour un déplacement donné, le choix du mode se fait aussi, dans une certaine mesure, en fonction des niveaux de service comparés des différentes alternatives (automobile, train ou avion).

A chacun des déplacements décrits dans l'enquête ont été associés les temps de parcours correspondants selon les différents modes, et les fréquences moyennes pour le train et l'avion.

Une approche économétrique (modèle logit) consiste à modéliser, pour chacun des catégories de déplacements précédemment définies, la répartition modale en fonction des caractéristiques des offres de transport existantes sur les différentes relations.

Ce document présente une étude en cours de réalisation à la Direction des Routes par Mlle Elisabeth MAURICE, étudiante en Magistère d'économétrie à l'université de Paris X - Nanterre sur le thème :

Direction
des Routes

**Modélisation du partage modal des déplacements
de personnes à longue distance.**

Sous-direction
de la planification
et du Budget

Téléphone
(1) 40.81.13.89

Elle est fondée sur une exploitation de l'enquête « transports et communications » de 1993-94, enrichie d'une description de l'offre de transport.

Télécopie
(1) 40.81.12.62

Le champ étudié est celui des déplacements de plus de 100 km des résidents en France réalisés à l'intérieur du territoire national (Corse exclue).

Bureau R/PB/P

Affaire suivie par
Gilles Dumartin

A ce stade, ce document comporte les parties suivantes :

- Une présentation de résultats généraux de l'enquête (statistiques descriptives du partage modal notamment) (p 3). Cette présentation n'est pas strictement indispensable à la compréhension des parties suivantes.
- Les résultats de différentes analyses de données réalisées sur les déplacements : analyse des correspondances multiples (p 7), segmentations (p 12).

Les segmentations ont pour but d'obtenir des classes de déplacements homogènes en terme de mode de transport utilisé. Elles ont été réalisées au moyen des variables qui sont apparues discriminantes de ce point de vue au cours des analyses précédentes d'une part (ACM, classification), et qui sont généralement mesurables d'autre part.

- La phase suivante (p 19) n'a pu être qu'esquissée à ce jour, et est en cours de réalisation. Elle consiste, par une approche économétrique à modéliser le partage modal sur chacune des classes de demande précédemment définies, en considérant comme variables explicatives les caractéristiques des différentes offres modales susceptibles d'avoir été utilisées. En pratique, l'utilisation de modèle Logit doit permettre d'évaluer la part de la route par segment de déplacements en fonction des caractéristiques comparées des trois offres modales.

Cette approche pourrait alors être utilisée pour **simuler les effets en terme de report modal d'une modification de l'offre de transport** (routière, ferroviaire ou aérienne) - en terme de temps de parcours ou de fréquence - sur la base **d'une connaissance des seuls flux routiers.**

Annexes :

- Comparaison des volumes de trafic issus de l'enquête Transport avec des statistiques de trafic (p 20).
- Méthode de description simplifiée des offres modales de transport alternatives correspondant aux déplacements de l'enquête (p 22). L'objectif est de pouvoir intégrer les variables d'offre dans la modélisation.

Par la suite, les parties figurant en italiques sont des précisions techniques dont la lecture n'est pas indispensable à la compréhension générale du document.

**

Principales caractéristiques des déplacements étudiés

I - Avertissement

Le travail présenté par la suite visant à modéliser les choix modaux pour les déplacements de longue distance est basé sur une exploitation d'un extrait de la base issue de l'enquête « transports et communications 1993-1994 » de l'INSEE.

Cette enquête permet de donner une image globale de l'ensemble des voyages¹ à plus de 100 km réalisés par les résidents en France de 6 ans et plus au cours d'une année. Après redressement, le nombre des déplacements qui constituent ces voyages est évalué à 725 millions (soit 13,6 déplacements par personnes de 6 ans et plus).

Pour les besoins de la modélisation, tous les déplacements recensés n'ont pas été pris en compte : ont notamment été éliminés, les déplacements totalement ou partiellement effectués hors des frontières françaises, ceux ayant pour origine ou destination la Corse (compte tenu de la spécificité de cette région relativement au choix modal), ainsi que les déplacements dont certaines caractéristiques importantes ne sont pas renseignées (par exemple le mode de transport utilisé).

Au total, l'équivalent de 588 millions de déplacements ont été pris en considération (soit 81 % des déplacements couverts par l'enquête²).

Préalablement au travail d'analyses de données proprement dit, il nous a paru intéressant de présenter quelques caractéristiques générales relatives à la structure de ces déplacements de longue distance.

Il s'agit notamment de fixer quelques ordres de grandeur. Il convient cependant de garder à l'esprit que pour la plupart, ces résultats ne portent pas sur la totalité du champ observé par l'enquête, qu'il peuvent donc différer des résultats proprement dits de l'enquête « transport et communications »³ et qu'il ne figurent ici qu'à titre indicatif.

¹ De manière générale, un voyage est constitué de l'ensemble des déplacements depuis le départ du domicile jusqu'à son retour.

² Pratiquement, l'enquête recense et décrit 41 774 déplacements dont 33 881 correspondent au champ étudié ici.

³ Ces résultats doivent être publiés par l'INSEE et le Service Economique et Statistique du Ministère.

II - Résultats généraux

Quelques ordres de grandeurs directement issus de l'enquête⁴ :

* **Les déplacements sont réalisés à près de 80 % pour motif personnel.** Environ 16 % des déplacements sont d'ordre professionnel.

* Environ un tiers des voyages sont réalisés au cours d'une même journée, un autre tiers comprend entre une et trois nuit(s), un troisième tiers dépassant cette durée.

Les voyages pour motif professionnel sont effectués pour plus de la moitié au cours d'une même journée.

* Il s'effectue environ deux fois plus de **départs de voyages les Vendredi et Samedi** que les Mardi, Mercredi, Jeudi.

* **Les retours effectués le Dimanche** sont presque 2 fois plus nombreux ceux du Lundi et presque 3 fois plus ceux des autres jours.

Autres résultats relatifs aux seuls déplacements considérés par la suite⁵ :

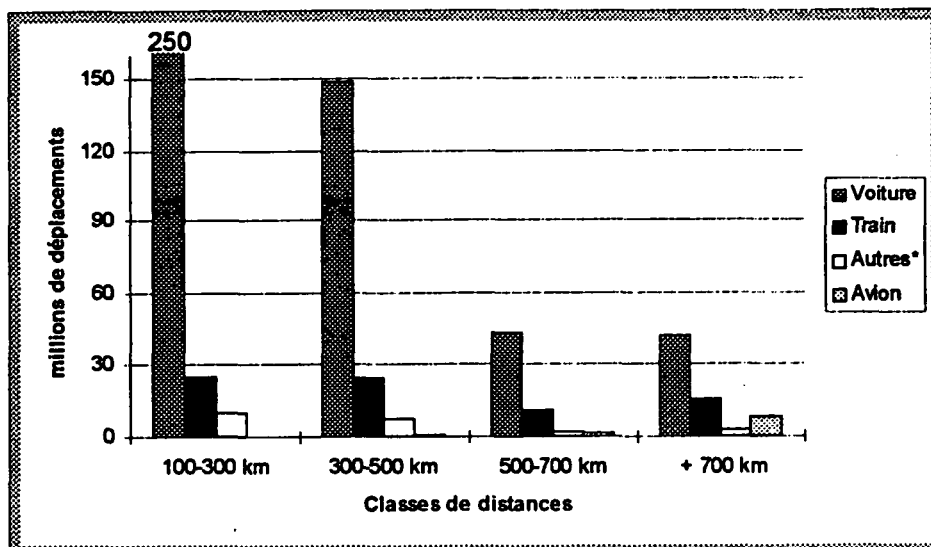
* Les déplacements intérieurs selon la distance et les modes utilisés

Les déplacements de moyenne distance sont majoritaires. L'usage de l'automobile est prédominant, même sur longues distances.

Nombre de déplacements par modes et classes de distances

⁴ portant sur la totalité du champ de l'enquête Transports et communications

⁵ Champ précisé en début de chapitre

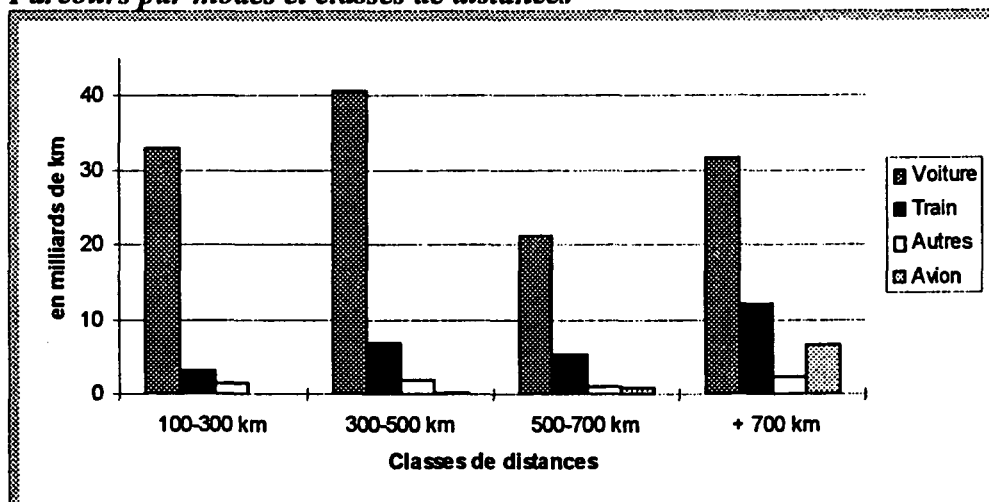


* Le mode « Autres » correspond essentiellement à l'autocar.

82 % des déplacements⁶ sont effectués en automobiles, moins de 13 % par le train et 1,8 % en avion.

Ramenée à la longueur des déplacements effectués (cf graphique ci-après), la part de l'automobile est de 75 %, celle du train de 16 % et celle de l'avion de 4,7 %.

Parcours par modes et classes de distances

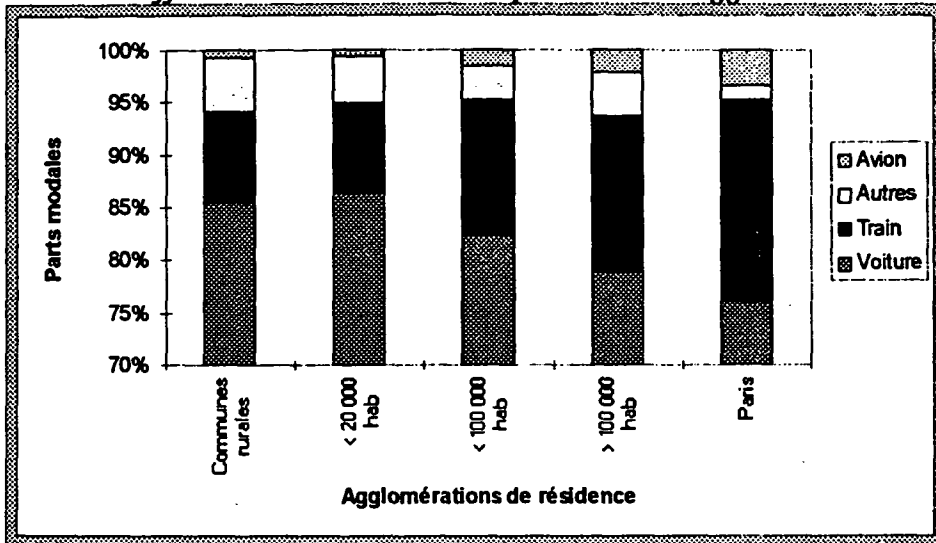


* Formes urbaines et parts des modes

Le graphique ci-dessous montre que l'attractivité du train et celle de l'avion sont relativement plus fortes pour les résidents des grandes villes qu'en milieu rural. Pour l'autocar, c'est dans une moindre mesure le contraire.

⁶ Un déplacement est relatif à une personne : plusieurs personnes se déplaçant dans une même voiture constituent autant de déplacements.

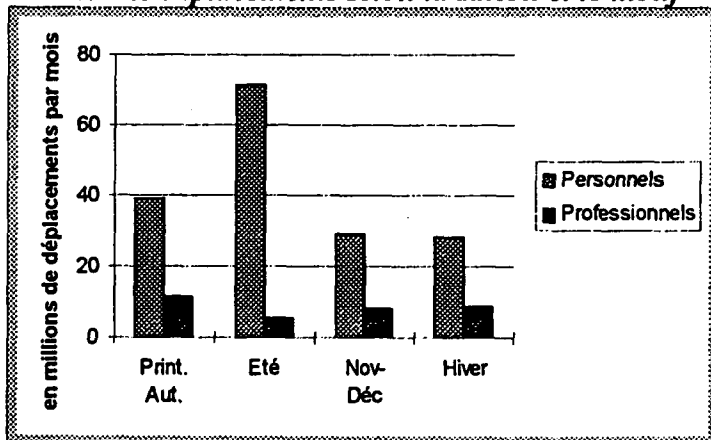
Parts des différents modes suivant l'importance de l'agglomération de résidence



*** Les fluctuations saisonnières de la demande**

Compte tenu de la prédominance des déplacements à motifs personnels, le nombre de déplacements effectués durant les mois d'Eté est deux fois supérieur à ceux des mois d'Automne et d'Hiver.

Nombre de déplacements selon la saison et le motif



**

L'analyse de données

Deux techniques d'analyse de données ont été successivement utilisées: l'analyse en composantes multiples et la segmentation.

I : L'analyse en composantes multiples

Principe :

Les unités statistiques considérées sont les déplacements : ils sont qualifiés par leurs caractéristiques propres et par celles des personnes les ayant effectués (descriptifs issus de l'enquête).

Les variables continues sont découpées en un nombre limité de modalités (ex : longueur des déplacements en 6 classes de distances).

Dans un espace géométrique, chaque variable correspond à un axe sur lequel se situent les modalités des déplacements. L'espace de départ comprend donc autant de dimensions que de variables (i. e. : de caractéristiques des déplacements) : Chaque déplacement y est repéré en fonction de ses modalités.

Une mesure de la 'distance'⁷ entre chacun d'eux est déterminée. Deux déplacements seront d'autant plus proches dans l'espace qu'ils ont un grand nombre de modalités communes ou voisines.

Dans cet espace, de nouveaux axes sont ensuite successivement déterminés (par combinaisons linéaires des axes originels), de telle sorte que :

- Une fois les déplacements projetés sur le premier de ces axes, les 'distances' mutuelles entre ces projections reflètent « au mieux » les 'distances' entre les déplacements eux-mêmes.

- Le second axe (perpendiculaire au premier) doit décrire lui aussi la plus grande partie de la dispersion des déplacements qui n'aura pas été traduite par le premier...

- ...et ainsi de suite pour les axes suivants.

Cette technique permet d'apprécier les liens entre les modalités des variables en les localisant dans un espace de faible dimension : celle-ci dépend de la part de l'inertie total du nuage de point que l'on souhaite faire apparaître.

Voir graphique ci-après.

Ici, seuls deux axes ont été conservés car ils permettent d'expliquer la majeure partie de l'inertie du nuage de déplacements initial.

Le premier axe (horizontal) est le plus explicatif : On peut considérer qu'il oppose clairement les déplacements à fortes valeurs du temps, à droite, à ceux à moindre valeur du temps.

D'un côté, on trouve en effets les déplacements pour motifs professionnels (et donc non effectués en famille), majoritairement effectués par les 25-49 ans, avec un niveau de diplôme plutôt supérieur à la moyenne.

A l'opposé, les déplacements sont effectués par des personnes souvent jeunes (- 25 ans), peu ou pas diplômées, majoritairement en famille et en été, pour des motifs personnels.

Bien qu'on ne puisse pas associer de manière très forte un mode de transport donné à chacun de ces deux types de déplacements, le train semble davantage utilisé pour le premier type de déplacements que pour le second.

⁷ distance figurant entre '' (i.e. : dans l'espace géométrique tel qu'il vient d'être défini), à ne pas confondre avec la notion de longueur de déplacement.

Le second axe fait apparaître clairement la relation entre longueur des déplacements (croissante de bas en haut) et modes de transport : voiture puis train et enfin avion pour la longue distance.

Sont aussi corrélés à la longueur des déplacements :

- le nombre de personne du ménage se déplaçant ensemble (qui diminue avec la distance);

- le niveau d'étude;

- l'importance de l'agglomération de résidence : les déplacements effectués par les Parisiens sont caractérisés par une longueur importante. A l'opposé, les déplacements plus courts sont proportionnellement davantage le fait de ruraux ou de résidents de petites villes.

- la motorisation des ménages : les déplacements de longue distance sont souvent le fait de personnes non motorisées (usagers du train et de l'avion) tandis que les multi-motorisés effectuent une plus forte proportion de déplacements plus courts.

Il semble à la lecture du graphique que ne posséder qu'une voiture pour un ménage amène à effectuer des déplacements plutôt plus long qu'en posséder plusieurs : c'est sans doute une interprétation simpliste car le nombre de voitures dépend aussi en partie de la taille des ménages. Une personne constituant à elle seule un ménage et possédant une voiture effectuera peut être plus volontiers des déplacements de longue distance qu'un ménage nombreux possédant 3 voitures...

Par ailleurs, il apparaît que les tranches d'âge supérieures (50-64 ans et a fortiori les 65 ans et plus) effectuent une forte proportion de déplacements à longue distance. Les plus âgés sont les moins motorisés et se déplacent en autocar.

Il apparaît aussi que les déplacements effectués au printemps et à l'automne sont de portée plus longue que ceux effectués en été : cela peut être lié au fait que les déplacements de la mi-saison sont moins effectués en famille, sont à caractère plus individuels (bien que pouvant faire partie de voyages organisés) et sont, de ce fait, souvent plus longs.

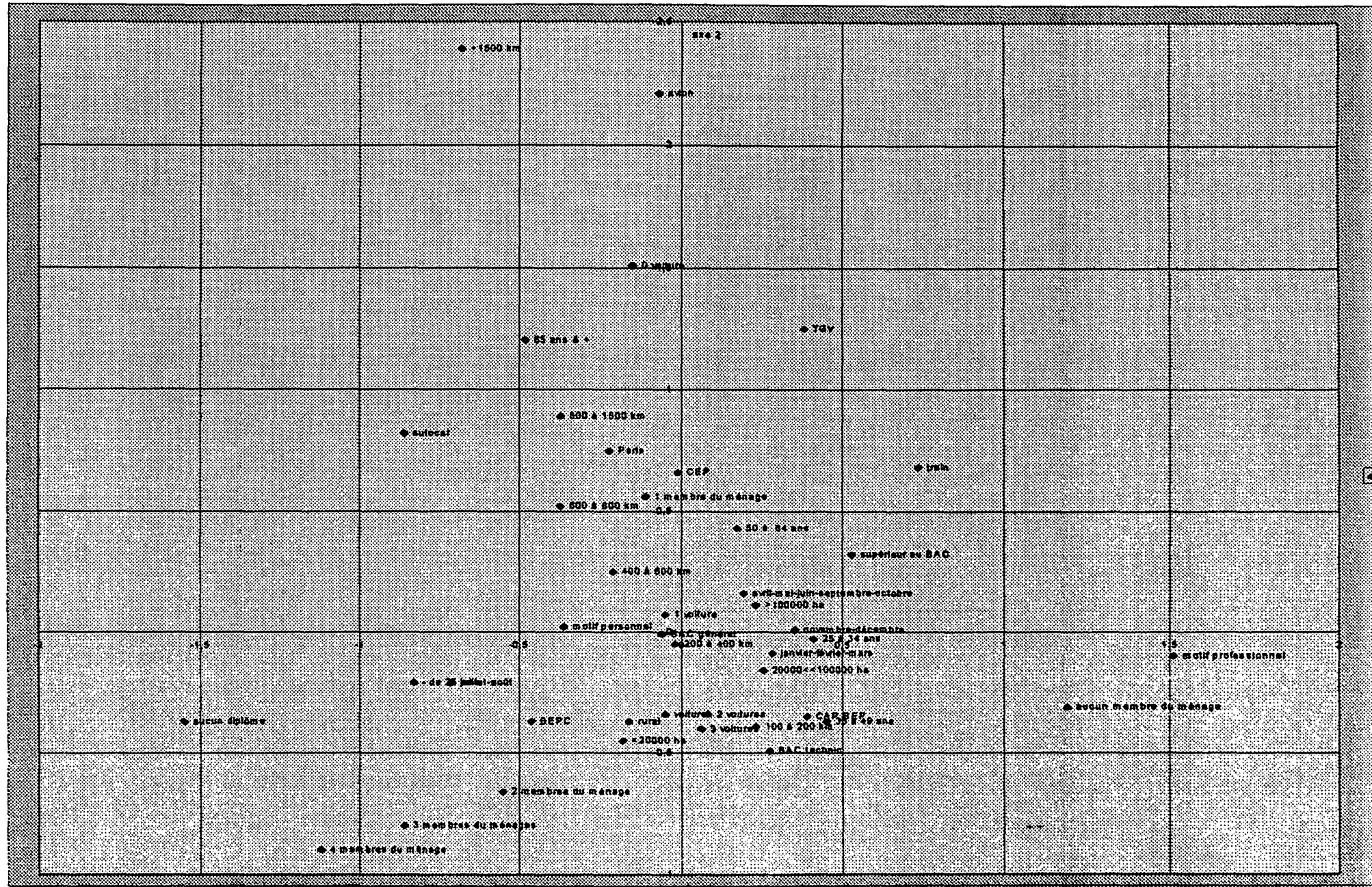
Synthèse :

- Déplacements de types personnels et professionnels se distinguent nettement.

- Les déplacements effectués en famille (comprenant au moins un enfant) sont plutôt de moyenne distance et essentiellement effectués en voiture. Ils sont souvent le fait de résidents de communes rurales ou de petites villes.

- Les catégories **les plus âgées** (65 ans et plus), ainsi que **les couples** et **les Parisiens** effectuent des déplacements en moyenne **plus longs** (plusieurs centaines de kilomètres), souvent **en transport collectif** (et notamment en autocar).

3/11



II : La segmentation

Principe

Cette technique ne permet pas, comme l'analyse en composantes multiples d'étudier simultanément les liaisons entre les différentes caractéristiques des déplacements.

Par contre elle permet de bâtir une typologie de ces déplacements, la plus discriminante possible par rapport à une caractéristique donnée, en l'occurrence le mode de transport utilisé.

En pratique, le programme recherche, parmi les variables descriptives des déplacements, à la fois :

- celle qui est la plus discriminante du point de vue du choix modal,*
- à quel niveau, parmi les différentes modalités de ladite variable, il convient d'effectuer la segmentation.*

Le programme considère alors les deux segments ainsi constitués et réitère sur chacun d'eux la procédure (détermination de la variable la plus discriminante et segmentation...).

L'arbre binaire ainsi constitué dépend des variables retenues en entrée pour qualifier les déplacements.

On s'est fixé comme contrainte que seules des variables couramment étudiées lors des recueils de données (enquêtes OD routières notamment) soient utilisées pour constituer la ou les typologie(s). Ce principe a notamment justifié de retenir pour la suite du travail (phase de modélisation) la première segmentation, que l'on a jugée plus exploitable que la seconde.

*** 1ère segmentation :**

Voir l'arbre binaire correspondant.

Le motif du déplacement (personnel ou professionnel) apparaît être le facteur le plus discriminant par rapport au mode de transport utilisé :

- 84 % des déplacements personnels sont effectués par la route pour seulement 10 % en train (4 % en autocar et 1% en avion).*
- 70 % des déplacements professionnels sont effectués en voiture et 25 % en train (et 5 % en avion).*

* Pour les déplacements professionnels, le facteur qui intervient ensuite en premier lieu dans le choix modal est celui de la distance : moins de 300 km, entre 300 km et 700 km et au-delà de 700 km étant les seuils les plus discriminants.

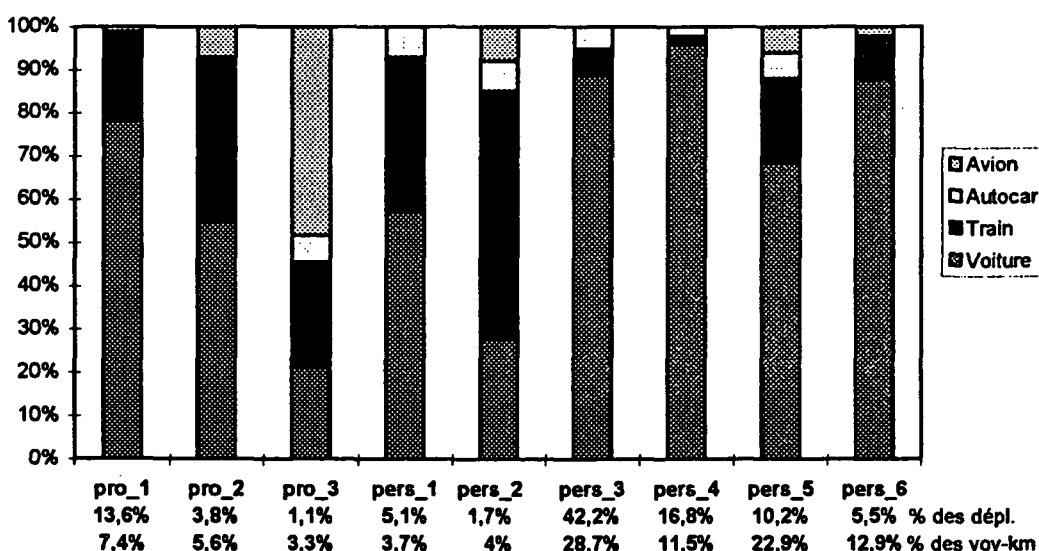
* Pour les déplacements personnels, le 1er déterminant du choix modal est la motorisation ou pas du ménage auquel appartient la personne se déplaçant.

- Pour les personnes non motorisés, le second facteur est celui de la distance (plus ou moins de 411 km).

- Pour les personnes non motorisées, le seuil de distance le plus discriminant en regard du choix modal est de 440 km (soit légèrement plus que ci-dessus). Vient ensuite la période à laquelle le déplacement est effectué : en été (Juillet-Août) ou en dehors de ces deux mois.

Si la technique de segmentation permet de construire une (ou des) typologie(s) des déplacements en regard du choix modal basée sur une discrimination par rapport au mode de transport utilisé, cette discrimination n'est que relative. Aussi est-il intéressant d'examiner, outre l'importance relative de chacune des classes de demande ainsi établies, la répartition des modes à l'intérieur de chacune d'entre elles.

Le diagramme ci-dessous illustre ces deux aspects :



Les deux séries de nombres correspondent aux parts relatives des classes dans l'ensemble des déplacements considérés, exprimés en nombre de déplacements et en voyageurs-km.

Définition des segments :

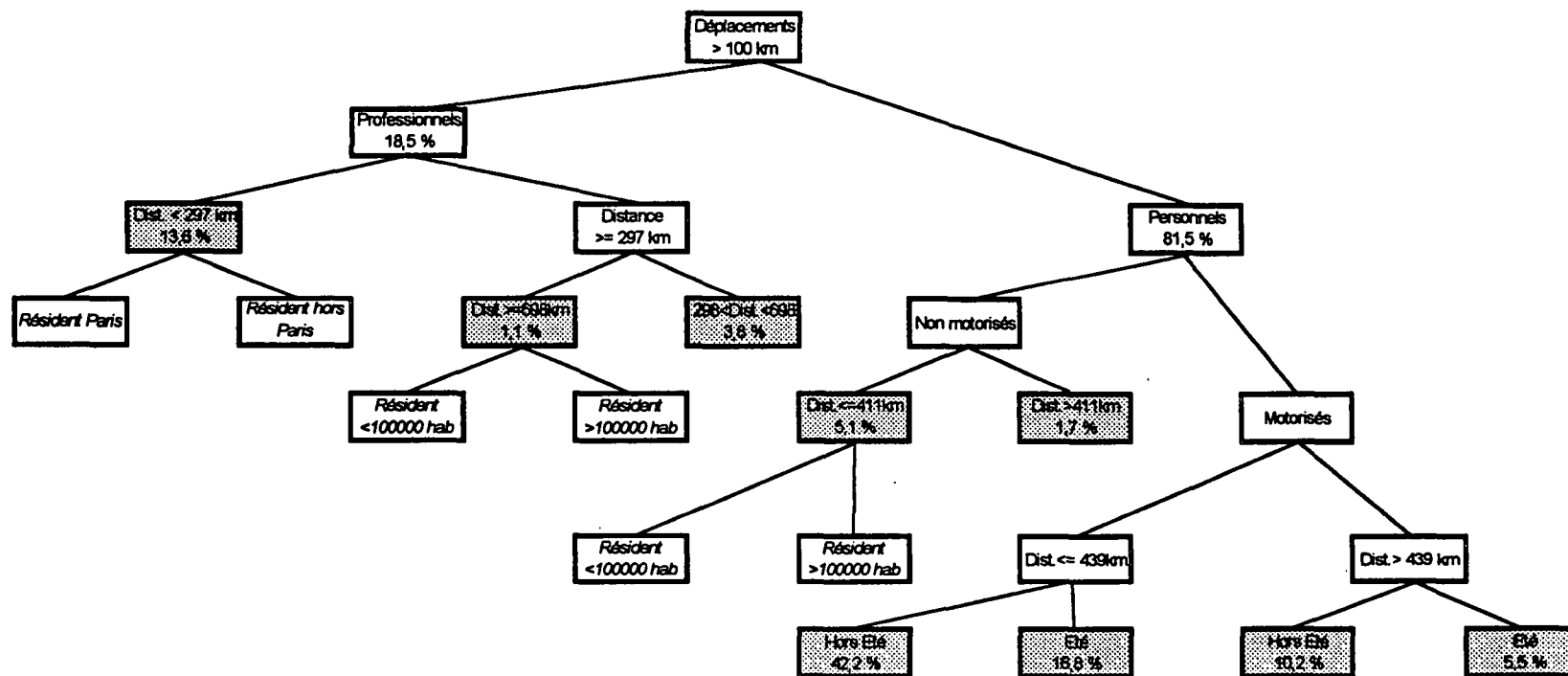
- pro_1 : déplacements professionnels de moins de 300 km
- pro_2 : déplacements professionnels compris entre 300 et 700 km
- pro_3 : déplacements professionnels de plus de 700 km
- pers_1 : dép. personnels effectués par une personne non motorisée de moins de 410 km
- pers_2 : dép. personnels effectués par une personne non motorisée de plus de 410 km
- pers_3 : dép. personnels effectués par une personne motorisée de moins de 440 km hors période de Juillet-Août
- pers_4 : dép. personnels effectués par une personne motorisée de moins de 440 km en Juillet-Août
- pers_5 : dép. personnels effectués par une personne motorisée de plus de 440 km hors période de Juillet-Août
- pers_6 : dép. personnels effectués par une personne motorisée de plus de 440 km en Juillet-Août

Près de la moitié des déplacements professionnels au-delà de 700 km sont effectués en avion.

Le train est utilisé pour les déplacements de « moyenne » distance pour les déplacements professionnels (entre 300 et 700 km) et surtout pour les déplacements personnels des ménages non motorisés au delà de 400 km (et dans une moindre mesure en-deçà de ce seuil).

L'usage de la voiture est évidemment prédominant pour les déplacements personnels des personnes motorisés (qui représentent 75 % des déplacements de longue distance).

Segmentation n° 1 - retenue pour la modélisation



Les segments finaux (sur lesquels portent la modélisation) figurent en gris - en % : part relative du segment par rapport à l'ensemble des déplacements - Des "sous-segments" figurent en italiques à titre indicatif.

3/5

*** 2ème segmentation :**

Voir l'arbre binaire correspondant.

Deux variables supplémentaires (la durée du voyage et l'effectif du groupe lors du déplacement) non retenues lors de la première qualifient les déplacements et donc peuvent intervenir dans la segmentation.

La variable « effectif du groupe » s'avère être d'ailleurs la plus discriminante par rapport au mode de transport utilisé : une première segmentation distingue les déplacements réalisés en groupes de 10 personnes et plus (se déplaçant majoritairement en autocar) des autres.

Parmi ces derniers, sont ensuite distingués ceux réalisés par une ou deux personne(s) de ceux réalisés par un groupe de 3 à 9 individus.

Les déplacements réalisés seuls ou à deux sont ensuite séparés suivant la longueur :

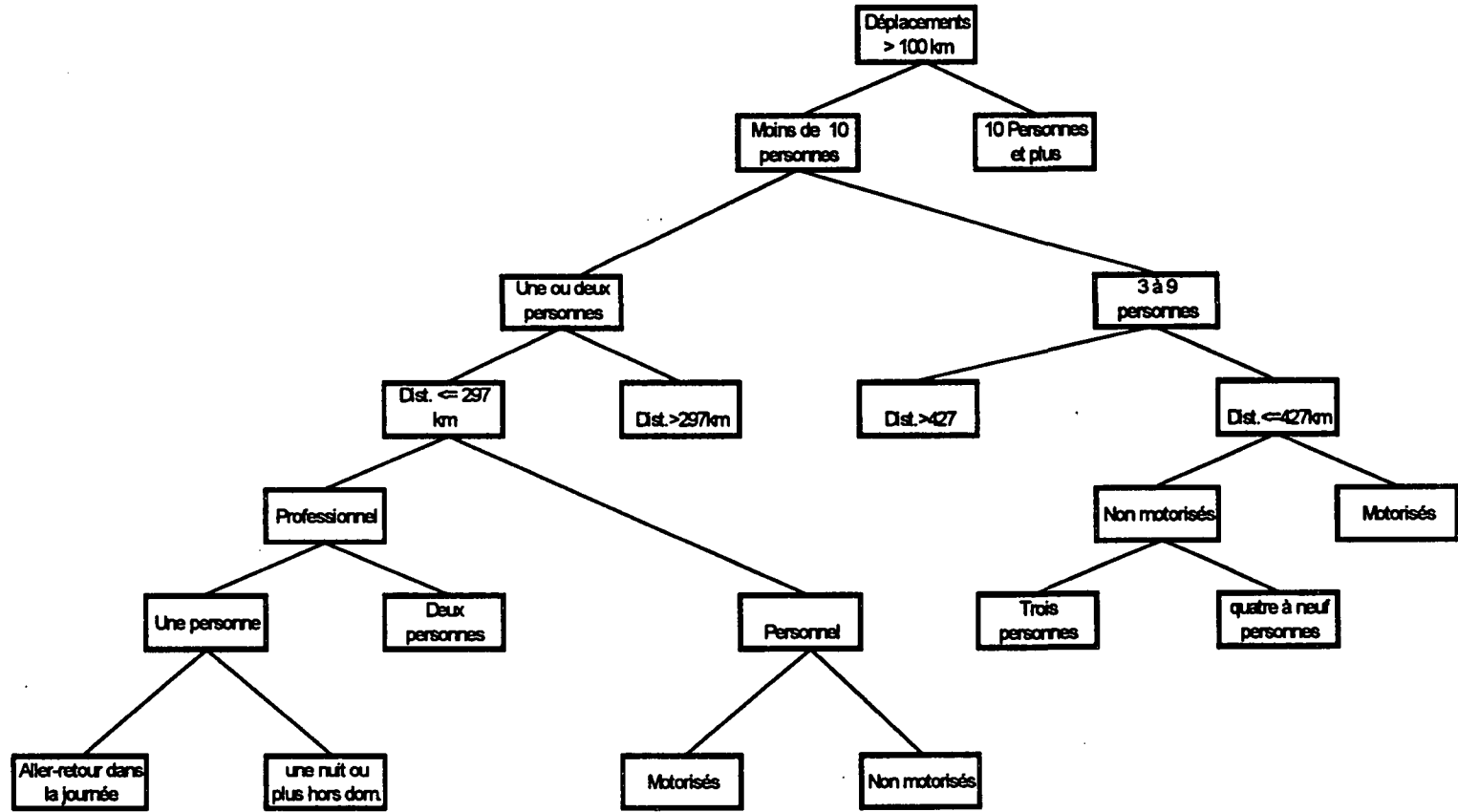
- Les déplacements de 300 km et plus constituent un « segment final ».
- Parmi les déplacements de moins de 300 km, c'est le motif, personnel ou professionnel, sur lequel est ensuite effectué la segmentation.

Pour les déplacements personnels, il apparait deux segments, distinguant les personnes motorisés d'une part et non motorisés d'autre part.

Pour les déplacements d'ordre professionnel (à moins de 300 km), la segmentation porte ensuite une nouvelle fois sur l'effectif : les personnes se déplaçant à deux ont tendance à se choisir proportionnellement davantage le train que celles se déplaçant seules (bien que dans tout les cas, le choix de la voiture soit prédominant).

Parmi les déplacements réalisés en groupe de 3 à 9 personnes, sont isolés ceux d'une portée supérieure à 430 km des autres. Le groupe des déplacements à moins de 430 km est ensuite scindé selon que la personne interrogée est motorisée ou pas.

Segmentation n° 2 - non retenue



317

La modélisation

Il s'agit, par une approche économétrique, de modéliser le partage modal sur chacune des classes de demande précédemment définies, en considérant comme variables explicatives les caractéristiques des différentes offres modales susceptibles d'avoir été utilisées. En pratique, l'utilisation de modèle Logit doit permettre d'évaluer la part de la route par segment de déplacements en fonction des caractéristiques comparées des trois offres modales.

Cette approche pourrait alors être utilisée pour **simuler les effets en terme de report modal d'une modification de l'offre de transport** (routière, ferroviaire ou aérienne) - en terme de temps de parcours ou de fréquence - sur la base **d'une connaissance des seuls flux routiers.**

Réalisation prévue : Automne 1997

Annexes 1 : Comparaison de quelques données de l'enquête avec des statistiques de trafic

Cette comparaison a un double but :

- Examiner la cohérence des trafics modaux globaux issus de l'enquête avec ceux fournis par les opérateurs.
- Evaluer ainsi la part du champ des déplacements étudiés ici relativement à l'ensemble des trafics.

1 : Les trafics automobiles

La comparaison des résultats de l'enquête « transports et communications » relatifs aux parcours automobiles avec les statistiques de circulation et l'estimation de la part des parcours automobiles ainsi couverte par l'enquête s'avère délicate.

* L'enquête évalue à 150 milliards de voyageurs-kilomètres les déplacements des résidents en France de plus de 6 ans réalisés en automobile dans le cadre de voyage à 100 km. La taille moyenne des groupes effectuant ces déplacements est d'après l'enquête de 2,9. En excluant les déplacements réalisés à 6 personnes ou plus (dont on peut penser qu'ils sont effectués avec plusieurs véhicules), le taux d'occupation est de 2,7 passagers par véhicules (ce qui reste élevé, compte-tenu notamment de la non prise en compte des enfants de moins de 6 ans). **Les parcours correspondants sont donc d'environ 56 milliards de véhicules-km.** Ces parcours sont essentiellement (mais pas totalement) effectués hors agglomération.

* D'après le rapport des « Comptes des transports de la nation » - année 1994 (p 104-105), 487 milliards de km ont été effectués en France en 1994, dont 151 en urbain, soit 31 %.

Les voitures immatriculées en France ont réalisé 352 milliards de km sur le territoire français, dont vraisemblablement plus de 31 % en ville (les véhicules étrangers, les poids lourds et véhicules utilitaires circulant davantage hors agglomération). En considérant une part de parcours urbains de 35 % (?), les parcours des automobiles françaises réalisés en France hors agglomérations serait de 229 milliards de km.

Le champ des déplacements automobiles considérés dans l'enquête correspond donc vraisemblablement au quart du total du trafic hors agglomération des automobiles françaises. (La raison première de cette écart étant la prédominance des déplacements de courte distance dans le trafic automobile.)

De surcroît, exclusion faite des déplacements internationaux et de ceux dont la voiture n'est pas le mode principal, le « champ automobile » considéré dans cette étude correspond à 127 milliards de voyageurs-km, soit de l'ordre de 20 % des déplacements automobiles effectués hors agglomération.

2 : Les trafics ferroviaires

* L'enquête évalue à 33,5 milliards de voyageurs-km les parcours ayant pour mode principal le train (y compris les parcours initiaux et terminaux non effectués par un autre mode).

Dans le champ étudié ici (faisant exclusion des déplacements internationaux), les parcours sont de 27,5 milliards de voyageurs-km.

* Selon les statistiques de la SNCF⁸, le trafic ferroviaire hors services régionaux d'Ile-de-France s'élève à 49 milliards de voyageurs-km et à 41,5 milliards de voyageurs-km si l'on exclut l'ensemble des services régionaux (essentiellement réalisés à moins de 100 km). La partie effectuée en France des parcours internationaux s'élève à 5,6 milliards de voyageurs-km.

Compte tenu des différences de champ entre les deux sources, leur comparaison s'avère hasardeuse mais il semble tout de même que l'enquête transports sous-évalue les parcours ferroviaires⁹.

3 : Les trafics aériens

* Le trafic aérien correspondant au champ étudié ici s'élève à 7,7 milliards de voyageurs-km. Il ne comprend notamment pas la partie des voyages à l'étranger effectuée sur une relation intérieure française (rabattement vers un aéroport international), ni les parcours effectués par des non résidents en France. D'après l'enquête transports 63 % des déplacements réalisés en avion sont internationaux.

* Le trafic aérien intérieur français s'élève en 1993-94 à 12 milliards de voyageurs-kilomètres (source DGAC). Il comprend en plus du champ considéré plus haut les « acheminements aériens » vers un aéroport international des déplacements à l'étranger ainsi que le trafic France continentale - Corse qui doit être de l'ordre de 1 milliard de voyageurs-kilomètres.

La comparaison de ces données ne fait pas apparaître d'incohérence patente. Dans ces conditions, le champ étudié ici représente à peu près deux tiers du trafic aérien intérieur français.

⁸ moyenne des années 1993 et 1994

⁹ L'enquête transports ne recense pas les déplacements des militaires et de certains étudiants.

parcours minimaux de zones à zones sur la base de la description du réseau français en 5000 arcs.

* L'offre ferroviaire est décrite à partir du fichier constitué par le Cesa (Centre d'Etudes Supérieures en Aménagement - Université de Tours), à partir de la base horaire SNCF du service hiver 1993-94. Ce fichier correspond à une matrice gare à gare (440 gares, ferroviaires ou routières, référencées au total), des meilleurs temps de parcours et du nombre de relations journalières (tenant compte des possibilités de correspondances et des temps d'attente en résultant).

* La base d'offre aérienne a été fournie par la DGAC (Direction Générale de l'Aviation Civile) : elle référence l'ensemble des services aériens commerciaux réguliers du printemps de 1995. Les données ont été synthétisées en terme de temps de parcours aérien et en nombre de relations quotidiennes pour chaque relation desservie.

- Les limites d'une telle description

Cette caractérisation de la qualité de service des différents modes en regard des déplacements décrits dans l'enquête présente deux types de limites :

- La première concerne la nature des variables d'offre : seule variable de temps de parcours pour les déplacements routiers, et, pour les liaisons ferroviaires (ou autocar SNCF) et aériennes, prise en compte, outre les temps de parcours de gare à gare (ou entre aéroports) des nombres de relations quotidiennes un jour de semaine hors été.

D'autres facteurs tels que les coûts de transport, la distribution horaire des services de transport collectifs, le nombre de correspondance ferroviaire ou l'accroissement de l'offre en été n'ont pas été pris en compte.

Différentes justifications à ces simplifications peuvent cependant être avancées :

i - La distribution horaire des services illustre la disponibilité du service au cours de la journée. Le nombre de relations constitue en partie un indicateur simplifié de cette notion.

ii - Les temps de parcours ferroviaires sont estimés en intégrant les temps de correspondance, seul le facteur pénibilité de ces correspondances étant occulté.

iii - Enfin le coût du transport pour l'utilisateur est une variable difficile à déterminer (coût réel \neq coût ressenti), fluctuant dans le temps et suivant les personnes. L'approche par segmentation de la demande doit permettre en principe de mieux tenir compte de manière implicite de telles fluctuations.

- La seconde question concerne la précision géographique de la description des offres. Dans l'enquête transports, les déplacements sont caractérisés par des origines et destinations communales. Il ne nous était possible pour des OD aussi précises, ni d'estimer

des temps routiers avec exactitude, ni de déterminer précisément les gares ou aéroports susceptibles d'avoir été utilisés et de décrire les parcours initiaux ou finaux vers ou depuis ceux-ci.

- La description de l'offre de zone d'emplois à zone d'emplois

Chaque origine et destination communales des déplacements de l'enquête transport ont donc été associées aux zones d'emplois auxquelles elles appartiennent. L'offre de transport pour chacun des trois modes est décrite entre chacune de ces zones d'emplois. L'erreur ainsi commise en se référant à un tel découpage peut être relativement élevée pour des déplacements de faible distance. Mais compte-tenu de la fréquente concentration des populations au centre des zones d'emplois, de nombreux déplacements doivent avoir réellement pour extrémités des centres de zones, ce qui peut justifier l'approximation.

* Les données routières ont été utilisées telles qu'elles ont été fournies par le Setra.

* Les temps et fréquences ferroviaires

On a cherché à relier chaque zone d'emplois à la gare principale qu'elle contient. Lorsque deux ou plusieurs gares se trouvent dans une même zone d'emplois, seule la gare considérée comme la plus attractive a été retenue. Lorsque aucune gare ne se trouve dans une zone d'emplois, il a été considéré que les déplacements ayant pour origine ou destination une commune de ces zones d'emplois ne pouvaient être réalisés par le train; (cf carte ci-jointe)

Aux déplacements effectués entre deux zones d'emplois « munies » d'une gare, il a été mis en correspondance les temps et nombre de relations entre ces gares.

Cette méthode revient à assimiler la « zone d'attractivité » de la gare à sa zone d'emplois.

Remarque 1 : L'agglomération parisienne a été traitée de manière spécifique. En effet, elle est constituée de plusieurs zones d'emplois mais peu d'entre elles ont une gare assurant des dessertes grandes lignes. La plupart du temps des rabattements vers les gares parisiennes sont nécessaires depuis les communes des zones d'emplois autour de Paris. La base d'offre ferroviaire à notre disposition ne décrit pas ces services de rabattements assurés par les réseaux de banlieue ou urbain.

Aussi avons nous décidé de redécouper l'agglomération parisienne en secteurs concentriques :

- Paris et la proche banlieue constituent une zone bénéficiant d'un accès direct aux gares parisiennes de grandes lignes.

- A l'intérieur d'une zone plus périphérique (desservie par les réseaux de banlieue, notamment RER), l'accès aux gares parisiennes nécessite un temps de rabattement supplémentaire, forfaitairement considéré de 1 heure.

- Au delà de cette couronne, la possibilité de venir prendre un train dans Paris n'a plus été prise en compte.

Remarque 2 : Un raffinement peut être apporté à la description de l'offre ferroviaire pour tenir compte de la proximité éventuelle des gares par rapport aux origine et destination communales des déplacements traités. Pour cela, à une variable binaire caractérisant les déplacements est affectée la valeur :

- 1 si, effectivement, dans les communes origine et destination du déplacement considéré il existe des gares, celles-ci étant de plus reliées par un service ferroviaire,

- 0 dans le cas contraire.

Une telle variable peut être ensuite introduite comme variable d'offre ferroviaire dans la modélisation du partage modal.

* Les temps et fréquences aériennes

Pour le transport aérien, il a été constitué « à la main » une carte des aires d'attractivité des aéroports correspondant à un regroupement de zones d'emplois (cf carte). L'étendue de ces aires sont fonction de l'importance présumée de l'aéroport correspondant. Selon ce principe, à une zone d'emplois ne peut correspondre qu'un seul aéroport.

Connaissant les services aériens existants entre ceux-ci, les déplacements de l'enquête « réalisables » en avion (c'est à dire dont l'origine et la destination sont à l'intérieur d'aires d'influence de deux aéroports reliés entre eux par un service aérien) ont été identifiés et caractérisés par le temps de vol et le nombre de relations aériennes.

Cette approche est assez simplificatrice dans la mesure où elle ne tient pas compte de la facilité d'accès à l'aéroport au sein d'une même aire d'attractivité de l'aéroport. En outre, contrairement à l'offre ferroviaire, il n'est pas tenu compte de possibles correspondances aériennes.

En définitive, les 33 881 déplacements de l'enquête étudiés sont caractérisés par un temps de trajet en automobile. 27 105 déplacements sont jugés réalisables en train et 9199 en avion, les temps et fréquences correspondants leur étant associés.

**

Questions relatives à l'exposé de G. Dumartin et E. Maurice (DR) :

P. Barette

Les analyses de segmentation ont l'air très efficaces. Il faut cependant être très prudent, ne pas oublier de variables par exemple et être attentif à la prédominance de certaines variables qui est parfois trompeuse.

G. Dumartin

Je suis d'accord; on a l'impression que dans certaines catégories on ne va avoir que des déplacements en train, alors que ce n'est souvent pas aussi discriminant en réalité.

P. Carle

Il faut également faire attention à la corrélation de certaines variables entre elles.

D. Lemaire

L'approche consistant à utiliser des techniques statistiques variées pour mieux appréhender les segmentations pertinentes vis à vis du choix du mode qui vont être par la suite introduites plus formellement dans des applicatifs de prévision pose le problème de savoir quels sont les paramètres que l'on doit prendre en compte. Les variables de l'enquête sont-elles suffisantes ou réductrices? Par exemple, les effectifs semblent pertinents, mais ne répondent pas à la question commerciale; il faut savoir qui sont les gens ensemble dans la voiture ou dans l'autocar. Si il faut entamer d'autres segmentations plus pertinentes, le problème va se poser de savoir quelles variables devront être appréhendées et introduites dans cette segmentation. Les impératifs commerciaux dans notre cas sont très clairs, et liés à des impératifs tarifaires ou à une incitation au voyage d'un autre type qui nous interpellent sur la définition des véritables variables.

G. Dumartin

Je pense que vous êtes là dans le domaine de la haute couture, et nous davantage dans le prêt-à-porter! Vous vous focalisez sur une infrastructure bien déterminée, et donc sur des marchés très pointus.

M. Houée

L'enquête transport permet bien sûr d'aller plus loin. La critique n'est donc pas forcément à appliquer à l'enquête dans ce cas précis.

C. Calzada

Nous avons en effet une connaissance très précise de la composition du groupe; il y a par contre un problème de données manquantes. La taille du groupe est par ailleurs très mal connue dans cette version de l'enquête transport 93 par rapport à celle de 82. A. Kerihuel et N. Gouider en ont parlé très rapidement dans leur exposé, notamment en ce qui concerne les effectifs liés aux autocars.

O. Morellet

Autre chose à propos de ces corrélations paradoxales entre taille du groupe et choix de la voiture: le fait d'avoir fait une segmentation sans introduire l'offre comme un des critères de segmentation me paraît regrettable dans la mesure où je suis persuadé qu'un certain nombre de variables apparaissent uniquement parce qu'elles sont très corrélées à l'offre. Il est évident que le transport collectif dépend énormément du type de zone origine ou destination, et le choix de la voiture également; il y a donc des corrélations qui sont dues à l'offre. D'autre part, pourquoi faire des LOGIT au sein de chaque segment? Le fait de segmenter implicitement par des caractéristiques d'offre va considérablement limiter la variabilité de l'offre à l'intérieur de chaque segment et donc probablement péjorer les possibilités de

calibrer un modèle d'offre. Le fait par exemple de segmenter implicitement en fonction des prix et des temps risque paradoxalement de ne pas faciliter la mise au point du LOGIT ou de tout autre modèle de choix modal en fonction des caractéristiques d'offre.

G. Dumartin

Nous y avons été à moitié sensibles; la segmentation des communes hors de Paris n'a pas été prise en compte, car cela découlait principalement pour nous d'un effet d'offre.

O. Morellet

Il me semble plus logique d'introduire ces caractéristiques d'offre dès le départ.

G. Dumartin

Nous avons pensé que c'était l'occasion de représenter deux marchés distincts et qu'il était plus facile de modéliser au sein de ces deux marchés distincts, plutôt que de prendre globalement le marché total et d'essayer de bâtir une équation qui couvre les deux.

**DEPLACEMENTS VOYAGEURS
INTER-REGIONAUX A LONGUE DISTANCE
MODELES D'INTERACTIONS SPATIALES
ET RECHERCHE DE PARTITIONNEMENT DE L'ESPACE**

**Christian CALZADA (SES/DEE)
Claude GRASLAND (Equipe P.A.R.I.S.)**

**Jean-Marc VINCENT (IMAG)
François OTTOGALLI (IMAG)**

Cette étude, qui associe des chercheurs de disciplines différentes visait un double objectif : l'estimation de modèle d'interactions spatiales classiques, puis la mesure d'éventuels effets de barrière ou anti-barrière de manière directe ou indirecte (recherche de partitionnement).

L'*estimation* d'un modèle à doubles contraintes pour les déplacements inter-régionaux de voyageurs à longue distance par mode emprunté et tous motifs confondus, est menée à partir de différentes sources matricielles (flux en nombre de voyageurs et en trafics) et pour différents types de spécification de la fonction de résistance.

Ces modèles sont ensuite dans un deuxième temps *généralisés*, en incluant de possibles effets d'appartenance (barrières ou anti-barrières géographiques), on teste ensuite l'hypothèse d'un espace français des transports spatialement homogène et isotrope. A cette fin une plate-forme informatique de recherche de partitionnement a été construite.

Les résultats présentés ne concernent que les meilleurs découpages de l'espace géographique, partitions de rang 1, ces résultats seront commentés en regard de plusieurs éléments: comparaisons pour un même mode entre matrices d'origine statistique différente, analogies avec cartes de potentiels de population et de densité de population en France métropolitaine en 1990.

DEPLACEMENTS VOYAGEURS
INTER-REGIONAUX A LONGUE DISTANCE

***MODELES D'INTERACTIONS SPATIALES
ET RECHERCHE DE PARTITIONNEMENT
DE L'ESPACE REGIONAL FRANCAIS***

- coopération entre chercheurs de disciplines différentes:
statisticien - géographe quantitatif - informaticiens
(SES/DEE - Equipe PARIS - LMC (IMAG))

- double objectif :
 - Estimation de M.I.S. classiques
 - * modèles à doubles contraintes

 - ** à partir de différentes sources matricielles ≠
(flux à longue distance en nombre de voyageurs et en trafics)

 - *** pour différents types de spécification
de la fonction de résistance

- Généralisation de ces modèles par :
 - * l'inclusion de possibles effets d'appartenance
(barrières ou anti-barrières géographiques)

 - ** test de l'hypothèse d'un espace français des
transports spatialement homogène et isotrope

- Construction d'une plate-forme informatique
de recherche de partitionnement

DONNEES UTILISEES

Les matrices de flux utilisées sont issues de plusieurs sources:

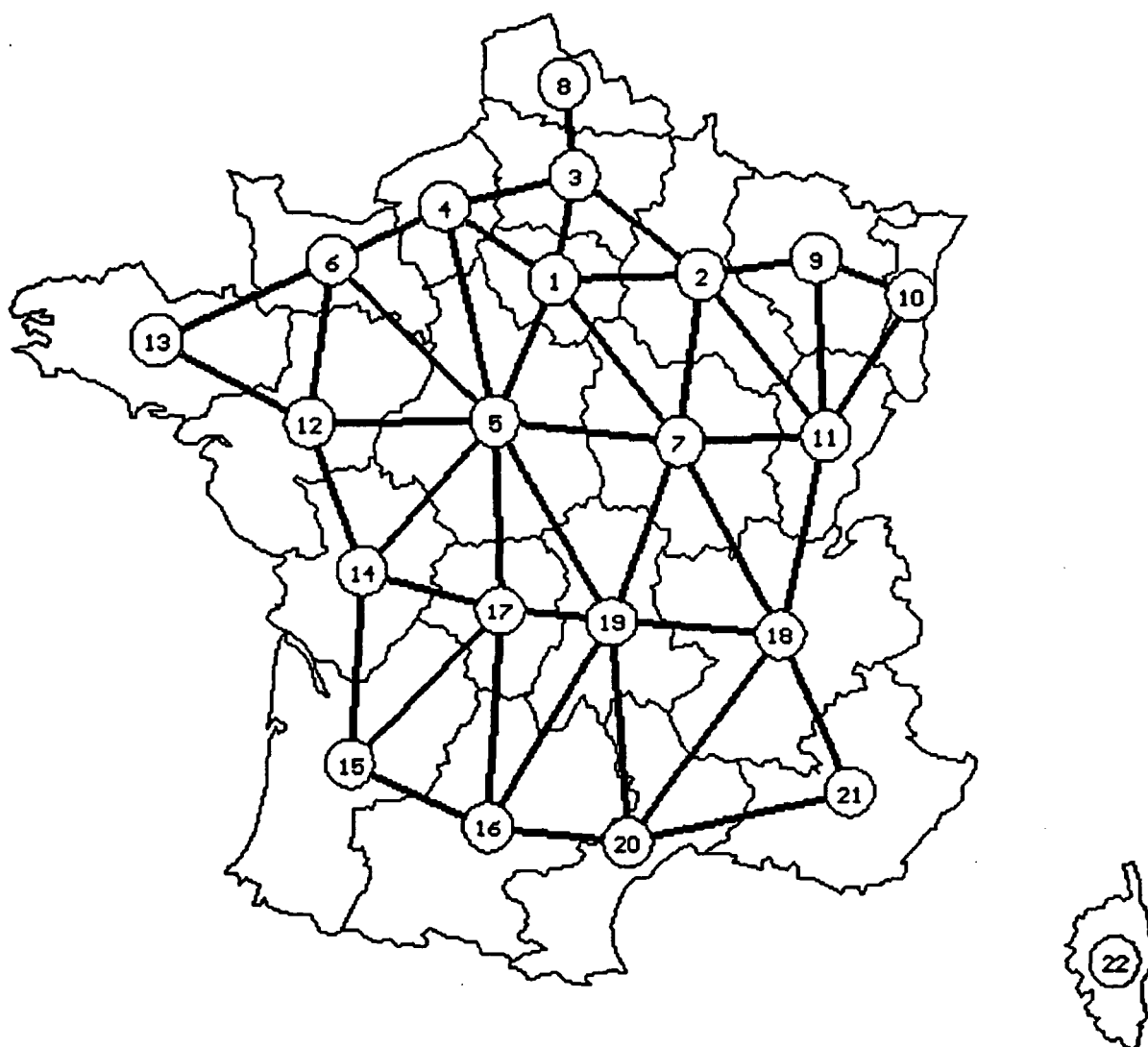
- de modèles de prévisions de trafic (modèle MATISSE) : matrices multimodales régionales de trafics en voyageurs et en nombre de véhicules pour le trafic français domestique en 1992
- matrices multimodales régionales issues de l'ETC 1993-94 bloc 'déplacements à longue distance'
- matrices de flux de véhicules VP interrégionaux (Setra) pour « une année 1993 » en TMJA, deux sens confondus
- ◆ 'symétrisation' des matrices régionales ($f_{ij} + f_{ji}$)
- ◆ retrait du mode aérien

9 types de matrices

- * 3 matrices issues de MATISSE: VP, Train et Total des flux
- * 3 matrices issues de l'ETC 93-94: VP, Train et Total des flux
- * 3 matrices routières en nombre de véhicules:
VP-SETRA (1ère et 2ème version), VP-MATISSE

ZONAGES

- ◆ A partir du fichier des zones d'emplois (INSEE)
- ◆ Topologie du graphe de contiguïté:
 - * exclusion de la région Corse
 - * hypothèse de non-rattachement des régions N°8 (Nord) et N°2 (Champagne)



EVALUATION DE L'INFLUENCE DE PARIS

L'influence de l'Ile de France a été décomposée en trois parties :

* cette région jouant un *rôle identique* à n'importe quelle autre région

* *région neutralisée* (PARIS NEUTRE), au sens où les flux peuvent transiter par la région mais sans s'arrêter

* *région déconnectée* des autres unités géographiques: aucun flux ne peut s'arrêter ni transiter par cette région

DISTANCES EMPLOYEES

◆ distance euclidienne (DEucl)

≈ distance '*à vol d'oiseau*' entre deux unités géographiques

$$DEucl = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

(x_i, y_i) : centre de gravité de l'unité i

(x_j, y_j) : centre de gravité de l'unité j

◆ construction de 3 nouvelles distances

prise en compte la population incluse
dans les zones d'emplois INSEE

\underline{D}_0 : distance euclidienne pondérée par
la proportion de la population des zones d'emplois

Soit les unités géographiques i et j et les zones d'emplois k
($k \in i$) et l ($l \in j$). La distance $D_{0i,j}$ est définie par:

$$D_{0i,j} = \sum_{k \in i} \sum_{l \in j} \frac{DEucl_{i,j} * pop[k] * pop[l]}{\sum_{s \in i} pop[s] * \sum_{t \in j} pop[t]}$$

avec

$pop_i[k]$: population de la zone d'emploi k
de l'unité géographique i

D₅₀ et D₈₀: distance euclidienne pondérée par la proportion de la population des zones d'emploi séparées d'au moins 50 km (resp. 80 km)

Soit les unités géographiques i et j et les zones d'emplois k (k ∈ i) et l (l ∈ j). La distance D₀i,j est définie par:

$$D_{50i,j} = \sum_{k \in i} \sum_{l \in j} \frac{DEucl_{k,l} * pop[k] * pop[l]}{\sum_{s \in i} pop[s] * \sum_{t \in j} pop[t]} * \delta [DEucl_{k,l} \geq 50]$$

avec

pop_i[k]: population de la zone d'emploi k de l'unité géographique i

$\delta [DEucl_{k,l} \geq 50]$: vaut 1 si $DEucl_{k,l} \geq 50$, 0 sinon.

(Idem écriture pour ≥ 80 km)

SELECTION DE LA MATRICE DE DISTANCE

- matrices de flux (9) x fonctions de distance (4)
x distances interrégionales (5)
= 180 configurations possibles !!!
- nécessité de choix sur l'une de ces contraintes = distance
sélection effectuée grâce au modèle à doubles contraintes
associé à la fonction de la distance de Pareto

L'analyse des erreurs présentée nous orientent vers le choix de
la distance D₀ comme distance la plus appropriée

SELECTION DE LA MATRICE DE DISTANCE

Modèle de Pareto

	DEucl	D0	D50	D80	D50+Prop	D80+Prop	MIN	MEDIANE
	Err	Err	Err	Err	Err	Err		
M31	0,0163	0,0080	0,0080	0,0587	0,0081	0,0724	0,0080	0,0122
M32	0,0197	0,0191	0,0194	0,0363	0,0189	0,0410	0,0189	0,0195
M34	0,0299	0,0279	0,0286	0,0709	0,0278	0,0843	0,0278	0,0292
MS15	0,0482	0,0507	0,0522	0,1675	0,0506	0,1978	0,0482	0,0515
MS16	0,0265	0,0247	0,0248	0,0301	0,0246	0,0316	0,0246	0,0256
MS19	0,0467	0,0499	0,0511	0,1303	0,0497	0,1517	0,0467	0,0505
VLS	0,0250	0,0105	0,0109	0,0981	0,0107	0,1232	0,0105	0,0180
VLS2	0,0163	0,0080	0,0080	0,0587	0,0081	0,0724	0,0080	0,0122
VLS3	0,0312	0,0250	0,0251	0,0770	0,0250	0,0598	0,0250	0,0282

333

SPECIFICATIONS

* Modèles d'Interactions Spatiales à doubles contraintes

$$F_{ij}^* = a_i O_i b_j D_j h(c_{ij})$$

$$a_i = \left[\sum_{i,j} b_j D_j h(c_{ij}) \right]^{-1}$$

$$b_j = \left[\sum_{i,j} a_i O_i h(c_{ij}) \right]^{-1}$$

O_i : somme des arrivées en i

D_j : somme des départs de j

$h(c_{ij})$: fonction *d'interaction spatiale (résistance)*

* Spécifications des modèles testés :

(I) modèle dit '*de référence*':

$$F_{ij}^* = a_i O_i b_j D_j h(c_{ij}) \quad \text{avec } h = \text{Id}$$

(II) modèle dit de *Pareto* (II):

$$F_{ij}^* = a_i O_i b_j D_j h(c_{ij}) \quad \text{avec } h(c_{ij}) = c_{ij}^{-\alpha}$$

(III) modèle dit de *Pareto+barrière*

$$F_{ij}^* = a_i O_i b_j D_j h(c_{ij}) \gamma_{ij}^A \quad \text{avec } h(c_{ij}) = c_{ij}^{-\alpha}$$

(IV): modèle dit de *barrière seule*

$$F_{ij}^* = a_i O_i b_j D_j \gamma_{ij}^A$$

$\gamma_{ij}^A / A_{ij} = 0$ (flux inter-régions), $A_{ij} = 1$ (flux intra-régions)

◆ MESURES DE QUALITE DES MODELES

Les erreurs de chaque type de modèle seront exprimées par rapport au modèle dit de référence (I) :

M¹. mesure en valeur absolue:

$$\text{Err}(F, F^*) = \sum_j F_{ij} - F_{ij}^*$$

M². mesure du carré de l'erreur:

$$\text{Err}(F, F^*) = \sum_j (F_{ij} - F_{ij}^*)^2$$

M³. mesure du χ^2_{F, F^*} (critère du type CHI2: χ^2_{F, F^*})

$$\text{Err}(F, F^*) = \sum_j \frac{(F_{ij} - F_{ij}^*)^2}{F_{ij}^*}$$

◆ temps de calcul : $M^3 > M^2 > M^1$

◆ données homogènes $\rightarrow M^1$ et M^2

◆ données hétérogènes $\rightarrow M^3$

◆ version plate-forme expérimentale : 1^{ère} version de M^3

$$\text{MA}(\chi^2) = \sum_j \frac{(F_{ij} - \text{MA}(F))^2}{\text{MA}(F)}$$

$$\chi^2_{F, F^*} = \sum_j \frac{(F_{ij} - F_{ij}^*)^2}{F_{ij}^*}$$

$$\text{Err} = \chi^2_{F, F^*} / \text{MA}(\chi^2)$$

$$\text{MA}(F) = \frac{1}{n \cdot (n-1)} \sum_j F_{ij}$$

Moyenne arithmétique des valeurs
de la matrice des flux observés F

\rightarrow calcul de χ^2 et normalisation par un ' χ^2 moyen'

→ dans la plate-forme expérimentale :

calcul d'une erreur propre à la matrice de distance employée

calcul d'une erreur de référence

pour chaque matrice en utilisant le modèle à doubles contraintes avec la fonction d'identité: $\chi^2_{F,F*id}$

La formulation de l'erreur du modèle (ε) devient:

$$\chi^2_{F,F*id} = \sum_i \frac{\sum_j (F_{ij} - F^* Id_{ij})^2}{F^* Id_{ij}} \quad \varepsilon = \chi^2_{F,F*} / \chi^2_{F,F*id}$$

(ε) représente la fonction de coût représentative de chaque option de modélisation

◆ à titre de test deux nouvelles variables avaient été définies
(prop_{ab} , \underline{d}_{ab})

$$\bullet \text{prop}_{ab} = \left[\sum_{i \in a} \sum_{j \in b} P_i P_j S_{ij} \right] / \left[\sum_{i \in a} \sum_{j \in b} P_i P_j \right]$$

avec

$$S_{ij} = 0 \\ = 1$$

(si zones d'emploi sont entre elles à un seuil de distance d_{ij})
(proportion de paires d'OD qui vérifient la condition S_{ij})

$$* \underline{d}_{ab} = \left[\sum_{i \in a} \sum_{j \in b} P_i P_j S_{ij} d_{ij} \right] / \left[\sum_{i \in a} \sum_{j \in b} P_i P_j S_{ij} \right]$$

RECHERCHE DE PARTITIONNEMENT DE L'ESPACE GEOGRAPHIQUE

- Problématique générale:

Fixer une règle de mobilité f permettant de reconstituer une matrice de flux estimés $[F^*]$ à partir de E , R et D

E et R : vecteurs des sommes de départ et d'arrivée

D : matrice de coût généralisé moyen des déplacements

$$[F^*] = f(E, R, [D])$$

“ L'hypothèse centrale de ces modèles (MIS) est que la distance, c'est à dire la proximité spatiale mesurée en kilomètres, temps ou coûts a une influence déterminante sur la probabilité de relation entre les habitants ou groupes sociaux répartis sur un territoire ”

La modélisation de l'effet d'appartenance doit permettre de mettre en évidence *“ des formes d'intégration territoriale spécifique, autrement dit des portions de territoires dans lesquelles les habitants adoptent un comportement migratoire significativement différent de celui des habitants d'autres portions du territoire ”*

[Grasland 1996]

Question posée :

- l'introduction d'une ou plusieurs partitions territoriales a priori dans le modèle entraîne-t-elle un gain d'ajustement ?
- recherche de la partition optimale a posteriori qui garantisse le meilleur ajustement
 - ⇒ problèmes de partition d'un ensemble connexe
- Recherche de la configuration optimale parmi l'ensemble de toutes les configurations possibles
 - ⇒ démarche lourde et coûteuse en temps de calcul :

phase n°1 de parallélisation :

répartition des calculs sur un nombre de processeurs p

phase n°2 : application de méthodes de répartition de charges afin d'éviter les déséquilibres entre les processeurs

phase n°3 : Classification des différentes partitions par le biais de la fonction de coût associée au modèle

⇒ la partition optimale, est celle dont la valeur de la fonction de coût est minimale (ε)

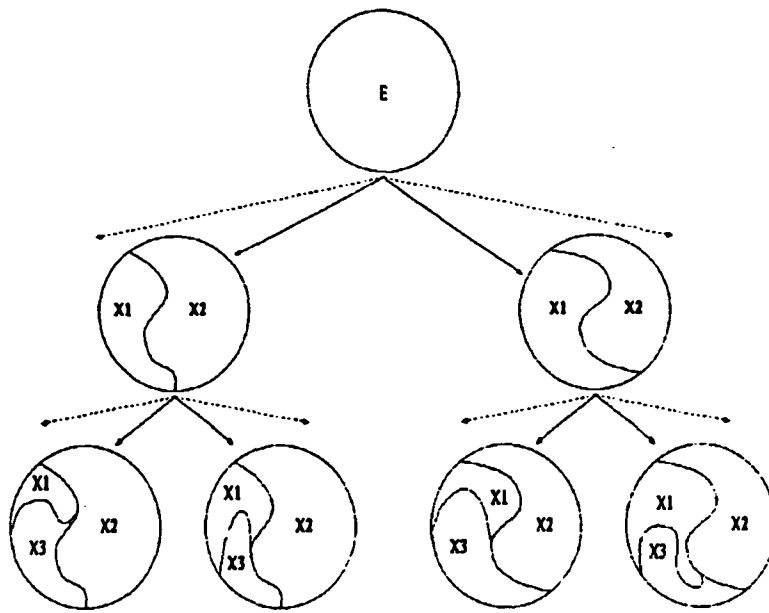
$$\varepsilon = \chi^2_{F,F^*} / \chi^2_{F,F^*id}$$

***Temps moyen de calcul par type de modèle utilisé
en secondes et heures/minutes/secondes***

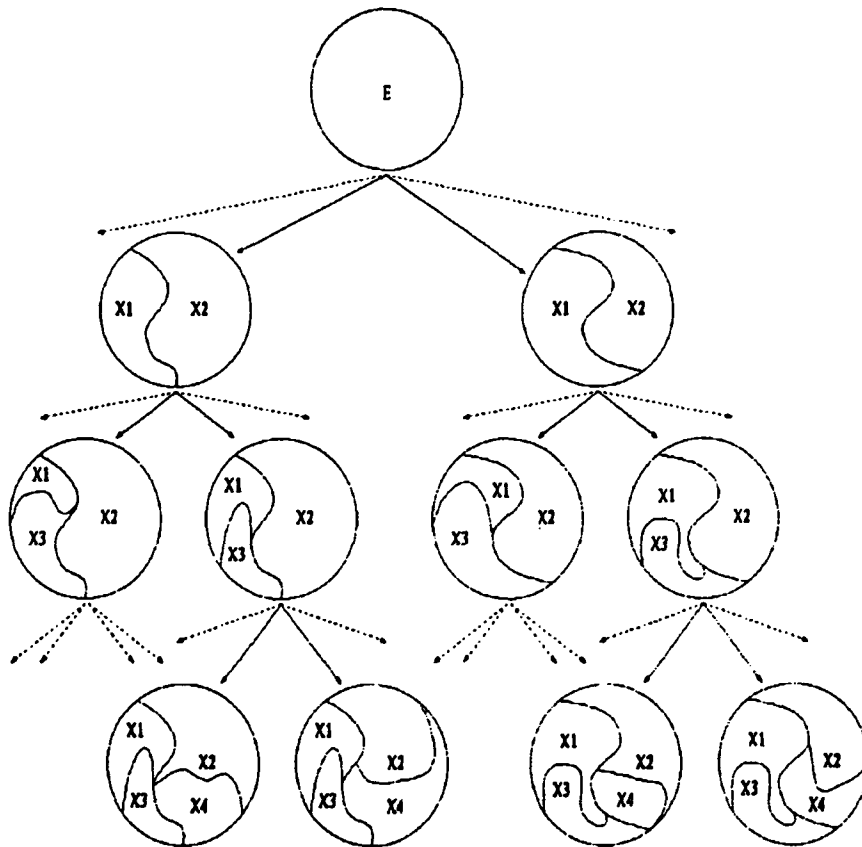
<i>Nombre de partitions en 2 parties connexes</i>				
	<i>Référence</i>	<i>avec PARIS</i>	<i>sans PARIS</i>	<i>PARIS neutre</i>
	1	2 085	1 116	3 103
BARRIERE	2	4 170 01:09:30	2 232 00:37:12	6 206 01:43:26
PARETO+BARRIERE	8	16 680 04:38:00	8 928 02:28:48	2 484 06:53:44

339

Partition en trois parties connexes



Partition en quatre parties connexes



Construction de k ($k > 2$) parties connexes

• ***Evaluation des paramètres du modèle***

⇒ fonction d'optimisation d' "Hooke" (f)
[Mark G. Johnson (déc.1994)]

'recherche directe' d'une solution

La méthode consiste à examiner un ensemble de solutions triviales afin de déterminer une stratégie de construction permettant de former un nouvel ensemble de solutions, inclus dans le premier, de meilleure qualité

Etape 0 : évaluer la valeur de f en un point initial x_0 un nouveau

Etape 1 : trouver x_1 tel que:

$$x_1 = x_0 \pm \delta_0 * f(x_1) < f(x_0)$$

δ_0 (pas de recherche de la solution) fixé à l'initialisation

$$\delta_0 = \delta_0 * \rho \text{ si } f(x_0 \pm \delta_0) \leq f(x_0) \quad \text{avec } \rho < 1$$

...

Etape n (condition d'arrêt) :

$$\text{si } f(x_{i+1}) - f(x_i) < \nu$$

ou lorsque le nombre d'itérations maximum est atteint.

Problèmes :

- comportement très sensible,
 - * aux conditions initiales (valeur de x_0 , δ_0 et ρ)
 - * à la forme de la fonctionnelle étudiée
- l'optima trouvé peut être local ou global sans qu'on puisse le savoir
- principal défaut de cette fonction est son caractère itératif qui la rend très coûteuse en temps de calcul

PRINCIPAUX RESULTATS

- meilleurs découpages de l'espace géographique, partitions de rang 1
 - le modèle simple de Pareto explique en moyenne plus de 80% des flux, avec une différence notable pour la matrice fer issue de l'ETC (56% de la variance expliquée)
 - l'introduction de l'effet de barrière dans le modèle de Pareto apparaît significative pour les matrices fer d'autant plus si l'on regarde les spécifications anti-barrière
 - la mesure des effets frontières (partition n°1) laisse apparaître une faible dispersion intermodale des valeurs
 - plate-forme informatique : support d'autres travaux
- ⇒ rôle des effets réseaux sur les flux de marchandises en Europe - réduction de l'effet frontière
approche dynamique des flux et réseaux

si l'on cumule l'ensemble des partitions de toutes les matrices par type de modèle et si l'on regarde les cartes dites de potentiel de population des analogies apparaissent immédiatement

Part expliquée par le modèle
(en référence au modèle dit de référence)

Types de modèles	REFERENCE	PARETO	IS SEULE		PARETO+BARRIERE		GAINS III/II	
	I	II	IV		III		en points	
			B	AB	B	AB	B	AB
	ERR	ε	ε	ε	ε	ε		
ROUTE [MATISSE]	0,355	93,8%	41,9%	3,6%	95,2%	95,2%	1,4%	1,4%
FER [MATISSE]	0,108	82,3%	-90,8%	-216,8%	86,1%	87,0%	3,8%	4,7%
TOTAL [MATISSE]	0,288	90,3%	40,3%	3,5%	93,1%	92,4%	2,7%	2,0%
ROUTE [ETC]	0,569	91,1%	39,5%	2,3%	93,5%	93,1%	2,4%	2,1%
FER [ETC]	0,055	55,5%	38,5%	9,5%	60,2%	63,8%	4,7%	8,3%
TOTAL [ETC]	0,422	88,2%	39,3%	2,2%	90,3%	90,8%	2,1%	2,6%
VLS	0,444	97,7%	41,9%	2,8%	98,4%	98,2%	0,8%	0,5%
VLS2	0,228	96,5%	45,5%	5,5%	96,9%	96,9%	0,4%	0,4%
VLS3	0,194	87,1%	43,8%	7,7%	88,7%	88,1%	1,5%	1,0%

442

Mesure de l'effet frontière

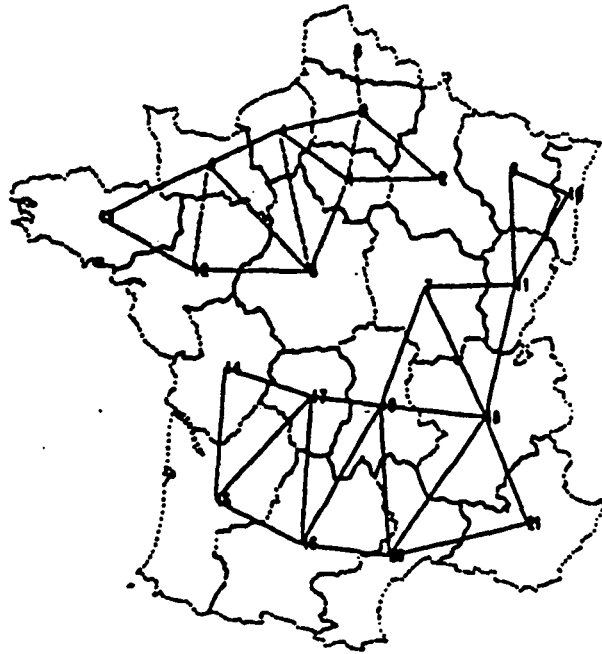
$\gamma > 1$

		-	PARIS NEUTRE	SANS PARIS
MATISSE ROUTE	M31	1,4	1,7	1,7
ETC ROUTE	MS15	1,6	1,7	1,7
MATISSE FER	M32	1,5	1,7	1,7
ETC FER	MS16	1,5	1,9	2,0
MATISSE TOTAL	M34	1,5	1,7	1,7
ETC TOTAL	MS19	1,4	1,8	1,8
SETRA ¹	VLS	1,8	1,8	1,8
SETRA ²	VLS2	1,4	1,4	1,4
MATISSE	VLS3	1,9	1,5	1,7

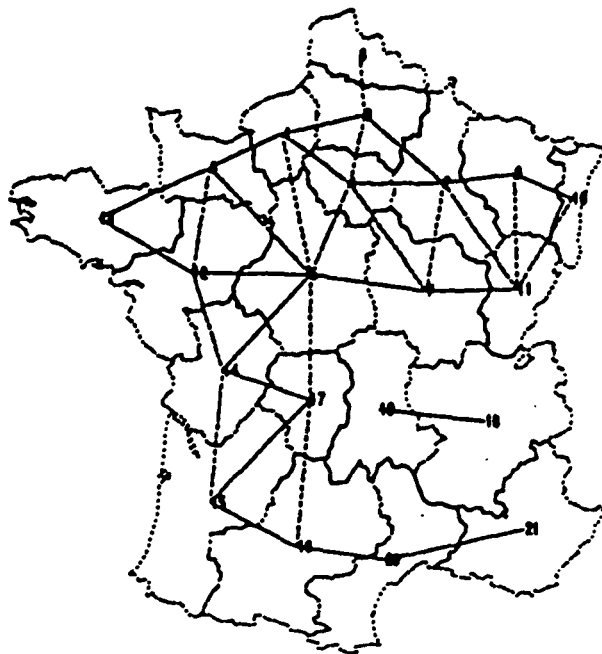
Spécification: **PARETO+BARRIERE**

ROUTE (véhicules)

MODELE (BARRIERE)

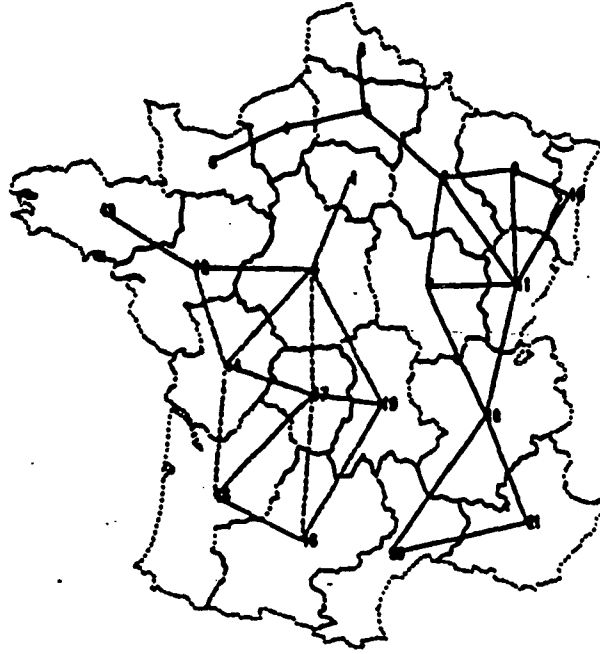


MODELE (PARETO + BARRIERE)

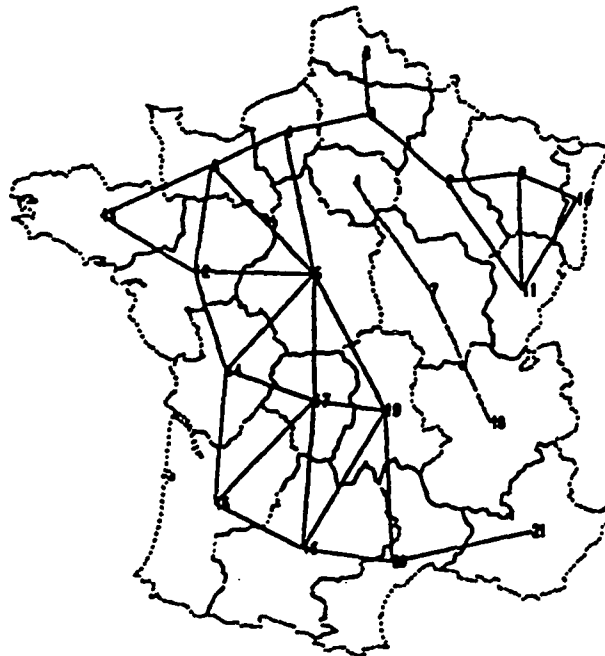


MODELE (PARETO + BARRIERE)

ETC ROUTE

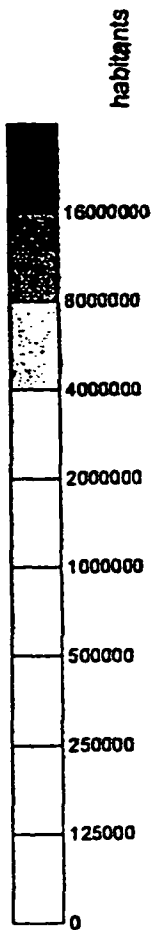
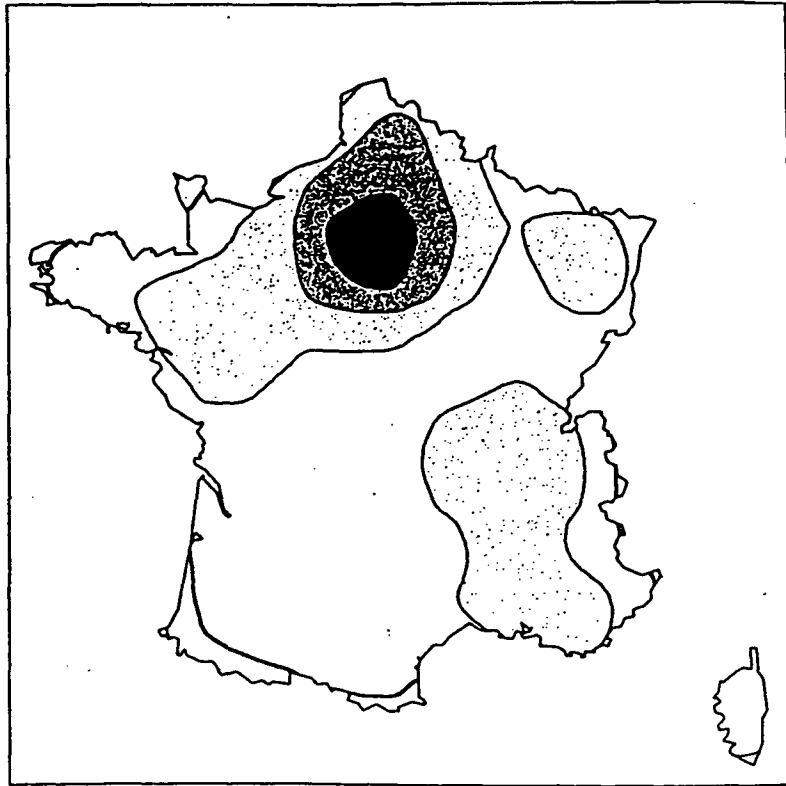


ETC FER



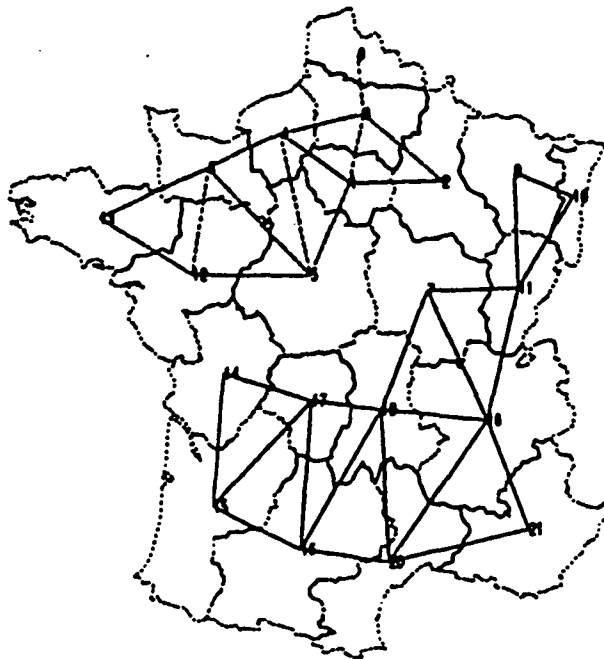
Potentiel de population en France (1990)

voisinage gaussien de portée 80 km



C. GRASLAND
CNRS-Equipe P.A.R.I.S.
(1997)

MODELE (BARRIERE)



Questions relatives à l'exposé de C.Calzada (SES) et C.Grasland (Université Paris I) :

D. Lemaire

Qu'est-ce que l'effet réseau ? Quand on parle de marchandises, on ne sait plus très bien ce que sont les réseaux, car ce qui est important, ce sont les services de A à B qui sont délivrés; en particulier, quel est le rôle de l'intermodalité pour un certain nombre d'acheminements vers l'Europe, par exemple en ce qui concerne les corridors de fret ? Les corridors de fret privilégiés sont-ils également soumis à cette problématique d'effet réseau ?

C. Grasland

Je pense qu'il s'agit d'un filtrage de l'effet distance. Quand on veut passer à l'identification de l'effet réseau, c'est une hypothèse de ce qui pourrait expliquer la barrière mise en évidence. Ce n'est pas évident, car il y a plusieurs facteurs qui interviennent dans l'équation de la barrière, l'effet réseau est donc difficile à déterminer.

C. Calzada

On peut soit introduire directement un effet réseau, soit regarder dans ce que l'on n'arrive pas à expliquer si en termes d'évolution des réseaux, il y a quelque chose à dire au niveau des résidus.

C. Grasland

L'effet de barrière n'a de sens qu'au cours du temps. La même matrice de flux reprise avec quelques changements dans les critères d'ajustement peut faire passer l'effet de barrière de 3 à 10. Il ne faut donc jamais interpréter les valeurs absolues d'un effet de barrière. On peut par contre spatialiser un effet de barrière, et savoir quelles régions subissent un effet de barrière important avec un autre pays, et donc arriver à une logique transport. L'approche est plutôt indirecte pour le moment.

D. Lemaire

Dans le cas du Transmanche, il y a eu deux enquêtes fondamentales O/D, l'une en 1991 qui traite de tous les modes avant l'existence du tunnel, et une bientôt disponible permettant de mettre en évidence les effets frontière.

M. Gaudry

Je ne suis pas du tout d'accord sur l'interprétation des effets frontières; on les trouve sur les flux commerciaux entre toutes les régions européennes, entre toutes les régions nord-américaines, à l'intérieur des provinces du Canada, entre les régions allemandes, etc...Ce sont des indicateurs d'absence d'intégration de marché, ils sont très utiles.

D. Lemaire

Je ne critique pas l'intérêt de ces coefficients mais l'utilisation de leur valeur précise. Je ne me permettrais simplement pas de comparer l'effet de barrière intérieur de deux pays, mais plutôt le fait, par exemple, qu'il a augmenté dans ces deux pays au cours des dernières années.

M. Houée

Vous soulignez surtout le risque de non comparabilité des différents chiffres, car cela est très dépendant du zonage entre autres.

Session 7 :

LES MODELISATIONS DE LA MOBILITE A LONGUE DISTANCE

* *Analyse prospective de la mobilité à longue distance*
O. MORELLET, P. MARCHAL, (INRETS/DEST)

* *Analyse du choix modal TGV-Nord sur l'axe IDF/NPC*
C. CALZADA (SES), H. HAMMADOU (CESURE)

* *Les modèles de concurrence air-fer en Europe :
de la concurrence à la complémentarité ?*
M. GAUDRY (CRT Montréal)

ANALYSE PROSPECTIVE DE LA MOBILITE A LONGUE DISTANCE

Exposé de O. Morellet et P. Marchal (INRETS/DEST) :

P. Marchal

Nous allons vous présenter une analyse de la longue distance, aussi bien à travers une interprétation des tendances sur la période 1975-1996 qu'une extrapolation de ces tendances à l'horizon de 2015. Nous tenons à associer deux personnes à cet analyse: D. Solleret également de l'INRETS ainsi que C. Calzada. Cette analyse s'appuie sur diverses sources de données statistiques avec entre autres les deux enquêtes les plus récentes: l'enquête transport 81-82 et l'enquête transport et communication 93-94, mais également sur des principes de modélisation développés au sein de l' INRETS. Nous essaierons d'avoir une approche chiffrée sur les effets du contexte socio-économique, avec le rapport entre la motorisation des ménages et l'évolution des déplacements, aussi bien en voiture particulière qu'en transport collectif. Nous verrons également tout ce qui touche à l'offre de transport et les différentes contributions que peuvent apporter les aménagements de l'offre de transport au trafic à longue distance.

Il faut d'abord préciser que cette analyse se veut multimodale, avec une approche chiffrée permettant de mettre en rapport les résultats avec certaines grandeurs connues d'activité des différents réseaux de transport à longue distance. On s'intéresse d'abord à des trajets effectués entre des lieux O/D; on voit ici un découpage en zones, celui qui a été retenu pour cette analyse. Ce choix est lié aux nombre d'années traitées puisque l'étude porte sur l'évolution de la longue distance entre 1975 et 2015. Nous avons par ailleurs à l'INRETS un découpage départemental qui intègre également le trafic international. Nous nous sommes également limités ici au trafic à plus de 100 km, pour ces mêmes raisons de description de l'offre de transport, et de validité de l'évolution des données au cours du temps. Le trafic considéré est donc celui effectué entre ces 42 zones.

Les différentes contributions dont je vous parlais sont l'évolution du contexte socio-économique (la croissance de la population, l'évolution des revenus et des modes de vie), l'évolution de la motorisation des ménages, et la diffusion du transport aérien, c'est à dire les effets de génération: nous savons que suivant l'âge d'une personne, celle-ci sera plus ou moins tentée de prendre les transports aériens. En ce qui concerne l'évolution de l'offre de transport du prix d'usage de la voiture, et des aménagements du réseau routier, nous verrons dans le cas du transport ferroviaire les modulations des tarifs, l'aménagement des dessertes. Pour l'avion, je parlerai de l'augmentation des fréquences et des modifications tarifaires.

Nous vous présentons donc des chiffres sur trois périodes principales: 1975-84, 1984-96 et une troisième période avec des hypothèses de planification socio-économique, dans un scénario médian pour la période 1992-2015. Il peut sembler étrange que la dernière période choisie commence avant la fin de la précédente : cela est du au fait que cette période est celle qui avait été considérée lors des exercices de planification au ministère des transports il y a un an, et que toutes les hypothèses étaient exprimées à partir de 1992. Nous avons donc des hypothèses avec des évolutions annuelles aussi bien pour le nombre de résidents, le nombre d'automobiles par habitant, et le prix du litre de carburant. Concernant les longueurs d'autoroutes à péage, pour la période future, elle est de 170 km par an; nous nous rapprochons ainsi de la période la plus récente 1984-96.

O. Morellet

Je vais à la fois parler de l'interprétation des tendances passées et de la justification de nos projections dans le futur. L'interprétation complète et les projections dans le futur ne sont possibles que si nous définissons un cadre théorique à partir des données d'observation, c'est à dire un modèle.

Nous avons utilisé le modèle MATISSE développé par l'INRETS, y compris dans le passé. Nous avons bien sûr vérifié que les valeurs du modèle étaient cohérentes avec les observations. Elles ne peuvent être identiques, du fait des conditions citées par P.Marchal, comme le découpage en zones. Il est très

probable que le trafic à longue distance croisse plus vite que le trafic à courte distance. De fait, les taux que nous allons annoncer sont supérieurs aux taux donnés par les indicateurs connus de trafic sur le réseau routier national et sur le réseau principal de la SNCF. Vous pouvez constater sur la courbe que l'on passe d'un taux de croissance de 3% par an pour les 20 dernières années à un taux de 2% par an pour les 20 années à venir. Si on compare cela soit à l'évolution passée de la consommation totale des ménages en volume, soit aux hypothèses d'évolution de cette consommation dans le futur, on constate que le trafic longue distance a augmenté avec un taux supérieur de 0,6 points à la consommation totale des ménages dans le passé, et dans le futur il devrait augmenter avec quelques dixièmes de point en moins. Il y a donc une inflexion, et l'un des objectifs de notre analyse est de comprendre à quoi elle est due.

Nous avons donc introduit la notion de contribution des différents facteurs de croissance du trafic en essayant de séparer les différents effets. En gros, chaque effet est obtenu en bloquant tous les autres facteurs sur des périodes assez importantes. La somme des contributions n'est donc pas égale à la contribution de la somme des facteurs. Cela permet de déterminer l'importance relative de chaque contribution par rapport aux autres. Pour chacune des contributions citées (population, revenus et modes de vie, motorisation des ménages, diffusion du transport aérien, prix d'usage de la voiture, aménagement du réseau routier, modulation des tarifs ferroviaires, aménagement des dessertes ferroviaires, fréquence des tarifs aériens), nous avons représenté 3 valeurs correspondant chacune à une période. Nous avons représenté la contribution en valeur absolue. Il s'agit donc de suppléments de trafic en voyageurs-km par rapport à une moyenne annuelle. J'ajouterais que le trafic français à longue distance est passé de 108 milliards de voyageurs-km en 1975 à presque 200 milliards en 1996. Le raisonnement me semble meilleur ici qu'en taux de croissance, car on oublie parfois qu'il s'agit d'une exponentielle discontinue.

En valeur absolue, on voit tout de suite si le phénomène est linéaire, ou très fortement exponentiel. Le facteur principal de croissance entre 1975 et 1996 est le facteur «population, revenu et mode de vie». Nous ne savons pas encore comment décomposer dans notre modèle ce facteur composite. Cet effet-là est quasiment linéaire d'année en année. Cela signifie que l'élasticité des ménages décroît dans le temps. On se trouve donc au point d'inflexion d'une courbe logistique classique. L'effet de la motorisation des ménages sur leur mobilité à longue distance est assez faible mais pas négligeable. Le prix d'usage de la voiture est un facteur plus important, il est essentiellement lié au prix de vente des carburants, et les trois périodes sont assez contrastées sur ce plan. L'aménagement des réseaux autoroutiers a une courbe exponentielle concave, et l'effet est de plus en plus faible, ce qui est logique. L'augmentation du nombre de vols et la baisse des tarifs a un effet non négligeable. Ce sont donc essentiellement les effets d'offre qui sont à l'origine du passage de 3% à 2%.

Je passe au trafic ferroviaire et aérien. Pour le train il y a une inflexion pour ce qui est de la longue distance: on passe de 1,3% en moyenne sur la période passée à 1,6 pour le futur. L'effet de la consommation reste stable, la diffusion du transport aérien n'existe pratiquement plus dans le futur, car l'apprentissage de ce mode de transport était important au début de la période passée, alors que dans le futur, il est absent. Il y a également moins d'effet négatif des autoroutes et des TGV, et cela contrebalance le moindre effet négatif de la diffusion du transport aérien et de l'aménagement des réseaux autoroutiers. L'effet total de la période future est donc plus fort que celui de la période passée, d'où ce léger accroissement du taux de croissance moyen annuel du trafic ferroviaire. L'hypothèse quant aux TGV qui seront construits d'ici 2015 est plutôt optimiste. Sans ces TGV, la croissance du trafic ferroviaire serait très fortement modifiée.

Je poursuis par l'avion: on passe d'un taux de croissance sur les 21 dernières années de 7,2% à un taux de 4,1% par an. Même si cela paraît être une chute, la courbe est quand même un peu plus que linéaire. On a donc quand même un taux de croissance supérieur à celui affiché par le ministère des transports. La contribution du facteur «population mode de vie» augmente, car l'avion est un moyen de transport cher. Ce phénomène de début de saturation de la mobilité concernait l'ensemble des individus, mais les

revenus étant de plus en plus élevés, de plus en plus de gens choisissent l'avion. L'effet des TGV est assez faible sur l'avion, alors que l'effet d'augmentation du nombre de vols et surtout de baisse du produit moyen aérien est assez fort et sera probablement du même ordre que ces dernières années. La baisse du produit moyen était auparavant due à l'augmentation de la gamme des réductions et non pas à la baisse du tarif plein. Nous pensons que cela sera différent dans le futur, du moins dans le cadre d'un scénario stable de libéralisation, car les compagnies seront obligées de baisser le tarif plein. Les élasticités du trafic aérien seront alors très différentes, car il y aurait alors probablement une récupération de trafic ferroviaire.

Je termine sur le partage du marché entre les différents moyens de transport au sein de la longue distance: en 1975 il était de 2/3 pour la voiture, 1/3 quasiment pour le train et très peu d'avion. En 96, l'augmentation de la part de marché de la voiture (en raison notamment de l'effet du prix des carburants) et de l'avion se fait au détriment du train. Si l'on se projette en 2015, on constate que la voiture perd à nouveau des parts de marché à cause d'un ralentissement des extensions autoroutières et surtout d'une évolution moins favorable du prix du carburant; l'avion continue à augmenter et le train se stabilise.

